A TEST-TAKING STRATEGY FOR EMOTIONALLY HANDICAPPED AND LEARNING DISABLED ADOLESCENTS

BY

CHARLES ALAN HUGHES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1985

Copyright 1985

by

Charles Alan Hughes

For my parents,

Life is enriched by choices and the freedom to make them. Thank you for a richer life.

ACKNOWLEDGEMENTS

The enclosed 100 or so pages culminate the most intense love-hate relationship I have experienced to date--my doctoral program. These acknowledgements are to those who have contributed to the love part.

To my committee goes my appreciation for their support and willingness to listen, answer, and advise. Each of you has provided a unique element to my program. Stuart, thank you for your exquisite professionalism and enthusiasm for special education. To the elder statesman of the department, Myron Cunningham, thank you for your calm wisdom and wonderful pictures which remind us that we are here for the kids. Cary, thank you for the early morning greetings and showing me that undergrads are people too. To the new "Doc on the Block" and fellow Irishman, Larry O'Shea, thanks for all the extra time and effort you gave and probably won't get credit for--we are Penn State. Ed, many thanks for coming in on short notice and providing a positive force on my committee.

Cecil, you get your own paragraph. I consider myself fortunate to have had you as my chair. You have provided me with support, consideration, encouragement, and opportunity as a professional and as a person. You have my lifelong gratitude.

I have been lucky enough to have had what I consider a second committee, which also deserves acknowledgement. To Leila Cantara goes my appreciation for excellent and insightful advice that has helped me through the maze. Our department lost one of its best advisors. Don Deshler and Jean Schumaker round out my second committee, and both of these individuals have given of their time and expertise to help me explore the area of my profession where my heart lies.

Some of my fellow docs deserve special acknowledgement because they not only made my program so much more enjoyable, but also made contributions to my dissertation. My thanks go to Pam Campbell and Lori Korinek for their expertise and willingness to help. Special thanks are given to Pam Hudson for sharing my interests and being my friend.

Carryout out experimental research in a middle school has not been easy. The logistical problems alone can drive one to, as Lee Clark is fond of saying, seriously consider managing a 7-11 store. Don Conn, ESE teacher at Lincoln Middle, deserves a huge acknowledgement for his continuous help in my study.

A special thank you goes to Henry Tenenbaum. Henry, you are truly one in a million. You have my respect and love for your expertise and friendship. I hope we remain friends for a long time.

My last acknowledgements go to my family: my parents, my sister, my daughter, and my wife. Mom and Dad, thanks for everything. Your love, kindness, and generosity have made all this possible and made

me a better parent. Sara, you have been a good friend. Thanks for laughing at my dumb jokes.

To my daughter, Hilary, go my love and respect. You have shown more patience with me than anyone and taught me that children are indeed the best teachers.

My wife, Kathy Ruhl, deserves my warmest regard for putting up with my occasional lack of grace under pressure. Kathy, thank you for your patience and invaluable assistance for these past four years. You are still the best teacher I have ever known.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT ix

CHAPTERS

I INTRODUCTION 1

Purpose 7
Questions 7
Definition of Terms 7
Delimitations 9
Limitations 9

II REVIEW OF THE LITERATURE 11

What Is Test-Wiseness? 11
Are Test Scores of Students Taught Test-Taking Skills Improved? 16
What Skills Should Be Included When Teaching Students Test-Taking Skills? 26
What Method Should Be Used to Teach Students Test-Taking Skills? 27
Why Should Mildly Handicapped (LD and EH) Students Be Taught a Test-Taking Strategy? 31
Summary 32

III METHOD 35

Subject Selection 35
Subject Description 37
Setting 37
Description of the Test-Taking Strategy 39
Instructional Procedures 43
Experimental Design 44
Reliability and Validity Procedures 50
Data Analysis 52

CHAPTERS

IV RESULTS 54

Summary of Results for Question One 54
Individual Data Summaries for Question Two 60
Summary of Results for Question Two 75
Summary 77

V DISCUSSION 79

Question 1: Do LD and EH Secondary Students Acquire a Test-Taking Strategy at 90 Percent Mastery When Taught in a Resource Setting? 80
Question 2: After Acquiring a Test-Taking Strategy at 90 Percent Mastery Level, Are Regular Classroom Test Scores of LD and EH Secondary Students Improved? . . . 83
Limitations 88
Future Directions 88

REFERENCES 90

APPENDICES

A INSTRUCTIONAL PROCEDURES 96

B METHODS FOR ESTIMATING ANSWERS 103

C KEY WORDS/PHRASES IN INSTRUCTION 104

D SPECIFIC DETERMINERS 105

E PRETEST 106

F ACQUISITION SCORES FOR ALL SUBJECTS 111

G GENERALIZATION SCORES FOR ALL SUBJECTS 112

BIOGRAPHICAL SKETCH 113

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A TEST-TAKING STRATEGY FOR EMOTIONALLY HANDICAPPED AND LEARNING DISABLED ADOLESCENTS

By

Charles A. Hughes

August, 1985

Chairman: Cecil D. Mercer
Major Department: Special Education

This study was conducted to determine whether middle school students labeled emotionally handicapped (EH) and learning disabled (LD) could learn a test-taking strategy and apply it successfully to mainstream classroom tests. Twelve subjects (six EH and six LD) were selected to participate. A multiple-baseline across subjects design was used to answer the experimental questions regarding acquisition and generalization. Scores on probes sampling test-taking behaviors were used to answer the first question of whether subjects could acquire a test-taking strategy at a 90 percent mastery level. Results show that all students with the exception of one did acquire the strategy. Additionally, subjects maintained acquisition over time albeit at a slightly lower level. Test scores in a mainstream class (i.e., science or social studies) were the data used to answer the second question of whether subjects who acquired the strategy improved

their scores. These data were collected over a seven month period and divided into two conditions. These conditions, baseline and during/after acquisition, were compared several ways. First, mean and median scores for each condition were calculated and then compared. Second, trend lines were calculated for both conditions and a comparison made. Third, means for the two conditions were converted to letter grades and changes between conditions noted. Results indicate that all subjects, except one, showed an increase in both mean and median scores between conditions. When the mean scores were converted to letter grades, 9 of the 12 subjects improved at least one letter grade. Analysis of trends indicated improvement in either slope or level for 8 of the 12 subjects. However, variability of test scores made visual inspection of data paths difficult. Generally, results indicated that subjects acquired and maintained the strategy and improved their test scores by doing so. Further replication with similar subjects is needed to strengthen conclusions about the efficacy of the test-taking strategy. Teachers of handicapped adolescents would be providing a valuable service to their students by teaching them test-taking skills which can help them succeed in mainstream settings.

CHAPTER I
INTRODUCTION

The majority of mildly handicapped, secondary students spend a portion of their school day in mainstream classrooms (Friend & McNutt, 1984; Heller, 1981) and, therefore, are expected to cope with the academic and social demands required in those settings. Due to learning characteristics of learning disabled (LD) and emotionally handicapped (EH) adolescents, successful functioning in the regular classroom is often difficult. For example, many LD and EH students exhibit below-grade level academic functioning and have difficulty following classroom rules (Deshler & Schumaker, in press; Kauffman, 1981). Additionally, evidence exists which suggests that expectations are the same for both handicapped and nonhandicapped students in secondary mainstream classrooms (Skrtic, 1980).

Descriptive data of both environmental expectations and learner characteristics of the mildly handicapped are available. Schumaker and Deshler (1983) summarized the literature on setting demand variables identified as critical to academic survival in regular secondary classrooms. These demands include gaining information from materials written at the secondary level, working independently with little feedback from the teacher, gaining

information from auditory sources such as lectures, participation in discussions, following class rules, and demonstrating knowledge by taking objective tests.

Research on characteristics of LD adolescents conducted through Kansas University, Institute for Research in Learning Disabilities (KU-IRLD) provides information as to how well the LD adolescent can meet demands of the secondary classroom. The findings of Warner, Schumaker, Alley, and Deshler (1980) indicate most LD adolescents do not possess adequate skills in gaining information from secondary level written information because of inadequate reading skills and lack of appropriate textbook usage strategies (e.g., scanning for specific information). Several KU-IRLD research reports indicate that, compared with nonhandicapped peers, LD adolescents appear to have greater difficulty working independently. For example, LD adolescents displayed lower rates of studying behavior, completion of homework assignments, requests for feedback, participate less in classroom discussion, and, when they do participate, emit half as many appropriate responses as nonhandicapped peers (Schumaker & Deshler, 1983). Schumaker and Deshler (1983) report violation of class rules by mildly handicapped students occurs at a higher rate and the rate increases as grade level increases. Finally, LD students seem to have difficulty demonstrating their knowledge. For example, they lack the skills for both studying for and taking classroom tests.

Unfortunately, learner characteristics of EH adolescents have not been as systematically researched as those of the LD adolescents.

However, information is available that suggests EH adolescents also exhibit behaviors hindering success in meeting secondary academic and social demands. EH students have been characterized by impulsivity and deficit independent work habits (Kauffman, 1981; Swift & Spivack, 1969). Kauffman (1981) states that EH students generally function below grade level as measured by standardized tests. Forness and Dvorak (1982) found EH adolescents have difficulty in testing situations.

Learner characteristics of EH and LD students and the demands placed on these students in regular, secondary settings create a situation which often results in nonproductive interaction. Consideration of this interaction leads to serious concerns about the ability of LD and EH adolescents to have a successful mainstream experience. However, examination of this interaction by the KU-IRLD provides a basis for an intervention approach to address the discrepancy between expectancies and skills.

The KU-IRLD developed a learning strategies (LS) approach based on epidemological findings which showed a mismatch between requirements for success in the secondary school and academic deficits of the LD student (Schumaker, Deshler, Alley, & Warner, 1983). This intervention approach was developed to teach students how to learn rather than teach them content (e.g., how to find information in a variety of texts as opposed to teaching information in a specific chapter). Alley and Deshler (1979) define LS as "techniques, principles, or rules that will facilitate the acquisition, manipulation, integration, storage and retrieval of information across situations and settings"

(p. 13). The intent of the LS approach is not only to teach students skills enabling them to meet classroom requirements, but also to teach them to generalize these skills to other types of tasks and settings. The KU-IRLD strategies are based on major curricular demands of the secondary school, and are organized into two strands which correspond to the types of student tasks. The two strands are strategies for gaining information from written and oral sources (e.g., textbooks and lectures) and strategies for expressing information as permanent products (e.g., written reports and tests) (Schumaker et al., 1983).

Results of studies indicate that while LD adolescents make little use of strategies before LS training, most can learn and successfully apply them after instruction. For example, LD adolescents were given a strategy designed to enable them to gain information from textbook chapters (Schumaker, Deshler, Denton, Alley, Clark, & Warner, 1981). Results showed the students learned the strategy and their grades on tests covering textbook material improved. Schumaker, Deshler, Nolan, Clark, Alley, and Warner (1981) report decreased error rates on written products (e.g., essays) for students taught an error monitoring procedure.

The standard for academic success in the mainstream classroom is mostly measured by the grade earned. According to Cuthbertson (1979), 60 percent of a student's total grade for a course is dependent on test scores. In a survey of secondary teachers Link (1980) found test-taking skills to be among the essential skills

for adequate performance in the classroom. Alley, Deshler, and Warner (1979) list poor test-taking strategies as one of the most common characteristics of LD adolescents. Havertape (1976) observed that LD students in test-taking situations often pay attention to the wrong part of the directions, are misled by irrelevant information, and are not persistent when searching for information. While information is not available on specific test-taking skills of EH adolescents, Forness and Dvorak (1982) described EH students as over-anxious and distractable during testing situations. Additionally, other writers have described EH students as having low impulse control and doing sloppy work (Kauffman, 1981; Swift & Spivack, 1969), both of which may hinder performance in a test-taking situation.

Despite documented importance of test scores and test-taking skills, there is evidence that most students, both regular and exceptional, do not receive direct instruction in test-taking strategies (Cuthbertson, 1979; Fluitt & Gifford, 1980). Cuthbertson (1979) found no consistent instruction offered by junior and senior high school teachers. Fluitt and Gifford (1980) report that none of the teacher preparation institutions surveyed prepared teachers to teach test-taking skills and most public, secondary schools surveyed did not systematically instruct students in this area.

That test-taking skills or strategies are not consistently or frequently taught to regular or exceptional students is unfortunate because test-taking skills can be successfully taught and can improve

performance on a variety of tests. Several investigators have shown that students who are test-wise[1] or consistently apply a test-taking strategy perform better on tests than non-test-wise peers (Crehan, Koehler, & Slakter, 1974; Eakins, Green, & Bushnell, 1976; Lee & Alley, 1981). Lee and Alley (1981) tested the efficacy of a test-taking strategy with LD adolescents. They report results of teaching a test-taking strategy, SCORER (Carman & Adams, 1972) to LD middle school students. Their findings indicate subjects did learn the strategy. However, due to design characteristics only limited conclusions can be drawn regarding how well SCORER generalizes to taking tests in the regular classroom. It then seems appropriate that low-performing, mildly handicapped students who are expected to meet demands of the regular classroom should receive instruction that would develop test-taking skills.

In summary, the majority of mildly handicapped secondary students are enrolled in mainstreamed settings and are expected to cope adequately with the academic demands placed on them. For the most part the criterion for success in the mainstream setting is grades received on tests designed and administered by the classroom teacher. There is evidence that the test-wiseness of a student does affect performance on tests and that test-wiseness can be improved by teaching test-taking strategies. Learning disabled and emotionally

---

[1]The term test-wise will be used in this paper to describe individuals who apply strategies to test-taking situations (e.g., time usage, attending to directions, and use of cues when guessing). The term is used due to its recurrent use in professional literature regarding test-taking skills.

handicapped adolescents exhibit behaviors which inhibit their ability to obtain test scores which fully reflect their abilities, skills, or knowledge. It can be assumed that LD and EH secondary students may benefit from instruction in a test-taking strategy designed to teach specific test-taking skills in a way which promotes generalization of the strategy across mainstream classroom testing situations.

Purpose

The purpose of this study was to evaluate the efficacy of teaching a test-taking strategy to LD and EH adolescents and to assess how well they generalize the strategy to the mainstream setting.

Questions

The study was designed to answer the following questions:

1. Do LD and EH secondary students acquire a test-taking strategy at 90 percent mastery level when taught in a resource class setting?

2. After acquiring a test-taking strategy at 90 percent mastery level, are regular classroom test scores of LD and EH secondary students improved?

Definition of Terms

Emotional Handicaps (EH). In this investigation, emotionally handicapped students are those students meeting the definition and eligibility requirements set forth by the Florida Department of Education for students identified as emotionally handicapped. That

definition is as follows: An emotionally handicapped student is one who after receiving supportive educational assistance and counseling services available to all students, still exhibits persistent and consistent severe behavioral disabilities which consequently disrupt the student's own learning process. This is the student whose inability to achieve adequate academic progress or satisfactory interpersonal relationships cannot be attributed primarily to physical, sensory, or intellectual deficits. The term does not include children who are socially maladjusted unless it is determined that they are emotionally handicapped (Florida Department of Education, 1981).

Learning Disability (LD). In this investigation, learning disabled students are those students meeting the definition and eligibility requirements set forth by the Florida Department of Education for students identified as learning disabled. That definition is as follows: One who exhibits a disorder in one or more of the basic psychological processes involved in understanding or in using spoken and written language. These may be manifested in disorders of listening, thinking, reading, talking, writing, spelling, or arithmetic. They do not include learning problems which are due primarily to visual, hearing, or motor handicaps, to mental retardation, to emotional disturbance, or to an environmental deprivation (Florida Department of Education, 1981).

Resource Classroom. The resource classroom is a type of service delivery system which provides specialized instruction for mildly handicapped students by a specialized teacher for a portion of the

school day. The type of resource class used in this study is called a varying exceptionalities resource room because all three mild handicaps (EH, LD, EMH) are served together (Deno, 1970).

Test-wiseness. A set of skills which enables a test taker to use the form and characteristics of tests and test situations to receive a score which reflects the abilities being sampled by the test (Lange, 1981).

Delimitations

The scope of the study was limited in several ways. The study was restricted to Lincoln Middle School, of the Alachua County Schools in Gainesville, Florida. Therefore, caution should be exercised in extrapolating results of the study to other geographic areas. Subjects used in this study were students from grades six to eight, who had been labeled emotionally handicapped or learning disabled according to Florida guidelines and were served in a resource room setting for part of the school day. Generalization of the results to students of other handicapping conditions is restricted. Students of similar handicaps and being served in a different level of the continuum of services (e.g., self-contained classes) were not included in the experimental population and caution is again advised in extrapolation of results.

Limitations

Alachua County Schools' policy of serving LD and EH students in a varying exceptionalities resource room may limit extrapolation

to similar students being served in categorical resource rooms. Additionally, the test-taking strategy used in this study is one of several and results based on this particular strategy should not be seen as reflecting the efficacy of other strategies.

CHAPTER II
REVIEW OF THE LITERATURE

The purpose of this chapter is to present a review of literature related to teaching students test-taking skills. For this purpose the review of the literature has been organized to address the following questions:

1. What is test-wiseness (TW)?
2. Do test scores of students taught test-taking skills improve?
3. What skills should be included when teaching students test-taking skills?
4. What method should be used to teach students test-taking skills?
5. Why should mildly handicapped (LD and EH) students be taught a test-taking strategy?

What Is Test-Wiseness?

Sidman and Willson-Morris (1974) conducted an experiment which led them to conclude that cues to correct answers may be present in test questions and/or the answer options. In this experiment adult subjects completed a 20-item, multiple-choice (MC) test on a topic with which they were unfamiliar. Typically, this test is given after

testees read a passage relating to the subject matter sampled by the test. However, for purpose of this experiment, the subjects were not given the passage to read, and yet 2/3 of the subjects answered at least 85 percent of the items correctly. This experiment clearly describes the concept of test-wiseness as defined by Millman, Bishop, and Ebel (1965). They define test-wiseness as "a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score" (p. 707). Furthermore, they state test-wiseness to be "logically independent of the knowledge of the subject matter for which the items are supposedly measures" (p. 707). Test-wiseness can therefore be seen as a set of cognitive abilities which can be applied to a variety of tests regardless of their content (Sarnacki, 1979).

Educational testing continues to be an issue in our society especially with regard to testing as a reliable measure of student learning or achievement (Gifford & Fluitt, 1978). Educational psychologists, concerned with testing and measurement of student ability, propose TW is a possible source of variance in test scores and consequently should be seen as a skill to be developed in students lacking the skill (Millman et al., 1965; Oakland, 1972; Rowley, 1974). It may be possible that test scores are effected by both too little and too much TW. McPhail (1980) suggests that more measurement error occurs when a student has too little TW than too much and TW should be taught in order to promote equality and opportunity for success for minority learners.

Millman et al. (1965) in a "classic" article analyzed TW for the purpose of providing a theoretical framework for empirical investigation. An outcome of their analysis was an outline of TW principles. Because of the comprehensiveness of the outline, frequency of use, and relevance for the present study it is included here (see Table 1).

The principles included in the first category of the outline, "elements independent of test constructor or test purpose" can be useful to the test-taker regardless of who constructed the test or the test's purpose. For example, the time use strategy in this category promotes efficiency in answering test items on any test having a time constraint. The second category, "elements dependent upon the test constructor or purpose," depends upon knowledge of the intent of the test constructor or the ability to use available cues resulting from poor test construction. For example, public school teachers lacking knowledge of sound test construction methods may overly qualify the correct option in a multiple choice test.

To summarize, test-taking is an important skill for students enrolled in public schools. Cuthbertson (1979) notes that junior and senior high school teachers administer tests after each unit of instruction. Additionally, students are required to take other tests such as minimum competency exams and standardized achievement tests. If, as has been pointed out, performance on tests is in part a function of TW, and performance differences are significant, it seems desirable to reduce these differences. As McPhail (1981) states, skills relating to TW should be taught to eliminate the unfair advantage

Table 1

An Outline of Test-Wiseness Principles

I. Elements independent of test constructor or test purpose.
   A. Time-using strategy.
      1. Begin to work as rapidly as possible with reasonable assurance of accuracy.
      2. Set up a schedule for progress through the test.
      3. Omit or guess at items (see I.C. and II.B.) which resist a quick response.
      4. Mark omitted items, or items which could use further consideration, to assure easy relocation.
      5. Use time remaining after completion of the test to reconsider answers.
   B. Error-avoidance strategy.
      1. Pay careful attention to directions, determining clearly the nature of the task and the intended basis for response.
      2. Pay careful attention to the items, determining clearly the nature of the question.
      3. Ask examiner for clarification when necessary, if it is permitted.
      4. Check all answers.
   C. Guessing strategy.
      1. Always guess if right answers only are scored.
      2. Always guess if the correction for guessing is less severe than a "correction for guessing" formula that gives an expected score of zero for random responding.
      3. Always guess even if the usual correction or a more severe penalty for guessing is employed, whenever elmination of options provides sufficient chance of profiting.
   D. Deductive reasoning strategy.
      1. Eliminate options which are known to be incorrect and choose from among the remaining options.
      2. Choose neither or both of two options which imply the correctness of each other.
      3. Choose neither or one (but not both) of two statements, one of which, if correct, would imply the incorrectness of the other.
      4. Restrict choice to those options which encompass all of two or more given statements known to be correct.
      5. Utilize relevant content information in other test items and options.

II. Elements dependent upon the test constructor or purpose.
   A. Intent consideration strategy.
      1. Interpret and answer questions in view of previous idiosyncratic emphases of the test constructor or in view of the test purpose.

Table 1--Continued

2. Answer items as the test constructor intended.
3. Adopt the level of sophistication that is expected.
4. Consider the relevance of specific detail.

B. Cue-using strategy.
   1. Recognize and make use of any consistent idiosyncracies of the test constructor which distinguish the correct answer from incorrect options.
      a. He makes it longer (shorter) than the incorrect options.
      b. He qualifies it more carefully, or makes it represent a higher degree of generalization.
      c. He includes more false (true) statement(s).
      d. He places it in certain physical positions among the options (such as in the middle).
      e. He places it in a certain logical position among an ordered set of options (such as the middle of the sequence).
      f. He includes (does not include) it among similar statements, or makes (does not make) it one of a pair of diametrically opposite statements.
      g. He composes (does not compose) it of familiar or stereotyped phraseology.
      h. He does not make it grammatically inconsistent with the stem.
   2. Consider the relevancy of specific detail when answering a given item.
   3. Recognize and make use of specific determiners.
   4. Recognize and make use of resemblances between the options and an aspect of the stem.
   5. Consider the subject matter and difficulty of neighboring items when interpreting and answering a given item.

(Millman et al., 1965, pp. 711-713)

some students may have in testing situations and therefore improve the validity of test results.

Are Test Scores of Students Taught Test-Taking Skills Improved?

To address the question of the efficacy of teaching students test-taking skills, data based articles on this topic were reviewed and are summarized in Table 2. The purpose of this section is to describe research characteristics of reviewed studies and to analyze their results in order to draw general conclusions about the effectiveness of test-taking programs.

The following criteria were used to select data-based research reports for this review of the literature:

1. Experimental question(s) examining the effect of instruction of test-taking skills on subject performance on teacher-made or standardized achievement tests.

2. Subjects in the study are school aged (kindergarten through 12th grades).

3. Subjects are attending public schools.

4. Content of the test-taking instruction focuses on general test-taking behaviors as opposed to coaching test content or the reduction of test anxiety.

The following sources were used in the search for appropriate publications: card catalogs of the University of Florida libraries, Current Index of Journals in Education, Exceptional Child Education Abstracts, Resources in Education, and reference sections of related articles or books. The search resulted in six research reports meeting the above criteria.

Table 2

Summary of Test-Taking Instruction Research

Study	Subjects	Independent Variable(s)	Dependent Measures	Results	Comments
Callenbach, 1973	48 students in the second grade.	Test-taking unit over a four week period; two sessions a week, 30 minutes each. (1,4,5)*	Stanford Reading Test. One form was administered immediately after the instructional unit, and a different form after four months.	Comparison of experimental and control groups on mean scores of both administrations of the Stanford Reading Test showed the experimental group scored significantly higher.	Use of a randomized posttest only design improves confidence of results. A follow-up assessment after four months supports the durability of the effect of the intervention.
Costar, 1980	290 students in the 4th grade.	Test-taking unit; Scoring High in Reading (Cohen & Foreman, 1978). Length of instruction not indicated. (3,5,6,7)*	Metropolitan Achievement Test. One form was administered after instruction, and a different form four months later.	Control group scored slightly higher on first administration and both groups scored approximately the same on the second administration.	The use of intact classrooms and the lack of randomization procedures for group selection cause concern about internal and external validity. No pretest was administered to account for any initial differences in reading ability between groups.

*NOTE: Refer to Table 3 for key to test-taking components used in reviewed studies.

Table 2--Continued

Study	Subjects	Independent Variable(s)	Dependent Measures	Results	Comments
Eakins, Green, & Bushnell, 1976	170 1st grade students in classes designed for low SES students.	Test-taking unit called The Behavior Analysis Test (BAT) which is designed to teach students how to take the Metropolitan Achievement Test. Experimental groups were either taught in a single presentation or multiple presentations. (1,2)*	Metropolitan Achievement Test (MAT administered after test-taking instruction.	Both experimental groups (Single and multiple presentation) scored significantly higher than the control group.	Methods of selection, design and analysis used were appropriate. Limiting the instruction to the format of the MAT reduces generalizability of results and lack of follow-up limits conclusions about durability of results.
Jongsma & Warshauer, 1975 Study I	91 5th grade students.	Test-taking unit called Test-Taking Tips. Unit took one and one-half hours over two days to teach. (1,2,4)*	Reading battery of the Stanford Achievement Test (SAT) administered two days after test-taking instruction.	No statistically significant difference between experimental and control groups on any subtests of the SAT.	While results were not statistically significant, the experimental group scored higher on all subtests than the control group. Comparison of the total score for the reading battery indicated the experimental group had a grade equivalent score of four months higher than the control.

Table 2--Continued

Study	Subjects	Independent Variable(s)	Dependent Measures	Results	Comments
Jongsma & Warshauer, 1975 Study II	54 5th grade students attending inner-city schools.	Same as Study I.	Comprehensive Test of Basic Skills (CTBS) administered two days after instruction.	No statistically significant difference between experimental and control groups on any subtests of the CTBS.	Again, while results were not statistically significant, the experimental group scored higher on all subtests and on total score. The experimental group ranged between three to five months higher than the control group on grade equivalency scores.
Lee & Alley, 1981	40 (grades 7-9) students labeled LD	Test-taking unit called SCORER (Carman & Adams, 1972). Unit covered 20 sessions, 30 minutes a day of small group instruction. (1,2,3,5,6)*	The pretest was an experimenter-made test covering information students had been exposed to in their regular classes within a month previous to intervention. The posttest measure was the subjects' score on a regular classroom science or social studies test.	All subjects acquired the use of the strategy as measured by practice materials. There was a significant difference favoring the experimental group on the posttest measure. Additional findings indicate that the higher the grade level the quicker subjects acquired and mastered the strategy.	While selection procedures and data analysis appear to be appropriate for the design of the study, there was only one opportunity to observe the effect of the strategy. The lack of on-going, long term measures leaves a question about durability of effect.

Research Characteristics and Results

Independent variables

Each of the six studies reviewed used a different program to teach students test-taking skills. Two investigations (Costar, 1980; Lee & Alley, 1981) used published strategies or techniques for test-taking while the remaining studies appear to use approaches constructed by the investigators. With the exception of one (Eakins, Green, & Bushnell, 1976), the studies used broad, general approaches to test-taking capable of application on a variety of objective test formats. Eakins et al. designed their instructional unit in a more limited fashion; familiarizing subjects to characteristics of the Metropolitan Achievement Test (MAT).

As shown in Table 3, all research reports include at least one of seven instructional components. Two frequently taught skills are efficient use of time and reading directions thoroughly; skills applicable to almost any type of test. The use of guessing strategies is also frequently used, however, this type of strategy is best used with objective tests (e.g., multiple choice or true-false) and is dependent upon test construction (Millman et al., 1965). The length of time spent instructing students varies. Time of instruction reported in the studies ranged from one hour (Eakins et al., 1976) to 20 sessions of 30 minutes each (Lee & Alley, 1981). Length of instruction or number of sessions is not reported by Costar (1980), and Eakins et al. (1976) do not quantify "multiple sessions." The variable of time spent on instruction and practice appears to be important. Bangert-Drown, Kulik, and Kulik (1983) report a strong

correlation between length of instruction and the size of the coaching effect.

Table 3

Components of Test-Taking Instruction and Frequency of Use in Studies From Table 2

Reference Number	Component	Frequency of Use
1	Efficient use of time	5
2	Read directions thoroughly	3
3	Skip and mark items unsure of	2
4	Mark answers appropriately	2
5	Use guessing strategies using cues	3
6	Review your work (e.g., answer all questions; don't change answers)	2
7	Read all answer options before marking an answer	1

Overall, the reviewed studies adequately describe test-taking programs used, or cited references where detailed descriptions could be found. Two investigators (Costar, 1980; Eakins et al., 1976) do not adequately describe the length of instruction which does appear to be a significant variable of teaching test taking.

Dependent measures

Of the six studies reviewed, five used a standardized achievement test as the dependent measure. These were Metropolitan Achievement Test (Costar, 1980; Eakins et al., 1976), Stanford Reading Test (Callenbach, 1973), Stanford Achievement Test (Jongsma & Warshauer, 1975), and the Comprehensive Test of Basic Skills (Jongsma & Warshauer, 1975). Only one investigation (Lee & Alley, 1981) examined the effect of teaching test-taking skills on the scores of classroom tests.

The use of achievement tests as dependent measures is warranted because all public school students take them and performance on these tests can have impact on the academic life of a student. That classroom tests were used in only one study does cause concern. Classroom tests occur more frequently and performance on these tests has more immediate impact than do standardized tests. As Cuthbertson (1979) indicates in a survey of secondary teachers, test scores make up approximately 60 percent of a student's grade and usually occur after each unit of instruction. It is curious that classroom tests are not used more often as dependent measures because there is some indication that teacher-made tests are more susceptible to TW than standardized tests (Sarnacki, 1979). This may be due, in part, to lack of training and expertise in writing tests by classroom teachers, which results in test items that are open to guessing strategies.

All measures used were objective tests, with the majority of items in multiple-choice form. The use of such objective tests is understandable for, as Smith (1982) points out, multiple choice is

the primary mode of mental measurement in the United States. Cuthbertson (1979) also found objective tests to be the most common form of assessment used by secondary teachers. Additionally, multiple-choice items appear to be vulnerable to TW. Smith (1982) and Rowley (1974) found TW to be positively correlated with performance on multiple-choice tests.

Characteristics of subjects

Grade level. The most frequently used subjects for investigations dealing with test-taking instruction were enrolled in the elementary grades. Approximately 94 percent (N=346) of students assigned to experimental groups were in grades one through five (excluding the third grade). One study (Lee & Alley, 1981) used secondary level students (grades seven through nine).

While the use of elementary subjects is justified, the limited use of secondary students is of concern. Secondary students are tested more frequently and are more able to learn a variety of test-taking techniques than younger students (Sarnacki, 1979; Slakter, Koehler, & Hampton, 1970). If, as several writers have found, TW is correlated with age, then the gap between secondary students who are test-wise and those who are not is even wider than at the elementary level.

School program. Students enrolled in regular education programs were the most frequently used subjects (96 percent). Included within this group were 170 students from low socioeconomic backgrounds who were served in classrooms with reduced enrollment (Eakins et al., 1976). Additionally, Jongsma and Warshauer (1975) used subjects attending an

inner-city school for one of their studies. Lee and Alley (1981) were the only researchers who used a population of exceptional children (LD) as subjects. These students were served in resource classrooms and attended regular content classes for part of the school day.

Only one study focusing on teaching test-taking skills used students enrolled in programs for exceptional students. When compared to normal learners, LD students appear to be at a greater disadvantage when taking tests. Quantitative differences in problem-solving skills and impulsive option selection are more characteristic of LD students than their normal peers (Havertape, 1976; Keogh, 1971). The special need for instruction in test-taking skills for LD students does not match the frequency with which it has been studied.

Results of research studies

In three of the six research studies included in this review significant differences are reported favoring the experimental group (Callenbach, 1973; Eakins et al., 1976; Lee & Alley, 1981). These results provide evidence that subjects taught test-taking skills obtain higher scores on objective tests than do students who do not receive training. In the two studies reported by Jongsma and Warshauer (1975) results were not statistically significant. However, the experimental groups scored higher on all subtests of the dependent measures used. The remaining study (Costar, 1980) resulted in a nonsignificant finding in which the control group scored higher on the dependent measure.

For investigations reporting positive or significant results, design characteristics are similar. For example, randomization procedures were used for subject selection and group placement in

each of these studies. The one study in which nonsignificant findings are reported used intact groups and yet made no attempt to establish equivalency between groups through a pretest. Additionally, analysis of variance procedures are used to analyze data in all of the studies with the exception of Lee and Alley (1981) who use analysis of covariance. Only one study (Callenbach, 1973) provides follow-up data. These data resulted from reassessment of subjects four months after initial measurement and found the experimental group again scored significantly higher. That five of six studies reviewed do not report follow-up data seriously limits conclusions about the durability of their results.

Summary of Research on Test-Taking Skills

Based on the results discussed above, there appears to be some data-based support for the efficacy of teaching students test-taking skills. The use as independent variables of a variety of skills (e.g., use of time, guessing strategies) in studies reporting positive results indicates that several types of test-taking behaviors may positively affect student performance on objective tests. However, little evidence exists on the long term effect of teaching students test-taking skills due to the lack of consistent inclusion of follow-up data in research reports.

All but one study used elementary school students enrolled in regular classrooms as subjects and used standardized achievement tests as the dependent measure. Lee and Alley (1981) are the only investigators using secondary level students classified LD as subjects

and a teacher-made test as the dependent variable. Therefore, little evidence exists supporting or not supporting training exceptional, secondary students in strategies to improve their scores on regular class tests.

What Skills Should Be Included When Teaching Students Test-Taking Skills?

Analysis of literature, both experimental (see Table 3) and nonexperimental (Carman & Adams, 1972; Fueyo, 1977; McPhail, 1981; Millman et al., 1965; Sarnacki, 1979), examining TW reveals several test-taking skills that consistently appear. They are

1. Strategies for efficient use of time such as working quickly, setting up a schedule for working through a test, and skipping and marking items unsure of.

2. Error-avoidance strategies such as reading directions carefully, reading all options before marking, and making sure answers are marked appropriately.

3. Deductive strategies such as marking out options known to be incorrect and choosing from the remaining options.

4. Strategies for making use of available cues such as choosing the most qualified option or not choosing answers that include specific determiners.

5. Reviewing the test form to ensure that all options are answered and marked correctly. Also, not changing answers unless they were marked incorrectly.

All of the above skills were used as a component of the independent variable in at least one of the experimental studies reviewed (see

Table 3). Unfortunately, a detailed analysis has not been done to factor out the effects of each skill. When choosing what skills should be included when teaching students how to take tests, it appears reasonable to include as many components as is feasible. It must be assumed then that all or some of these skills can be used by students to improve their test scores.

In keeping with the above rationale, the instructional unit on test-taking designed for use in this experiment includes all of the skills mentioned. The unit, PIRATES, is more thoroughly described in Chapter III.

What Method Should Be Used to Teach Students Test-Taking Skills?

In the previous section a variety of test-taking behaviors are identified which, when used, appear to increase scores on objective tests. These behaviors have been combined to form an instructional unit called PIRATES. The purpose of this section is to provide a rationale for a strategies approach to teaching students PIRATES and describe a specific set of instructional procedures designed to teach the strategy.

Rationale for a Learning Strategies Approach

Test-taking skills such as those included in PIRATES encompass behaviors which can be applied to a variety of objective tests and test settings. This is in contrast to teaching skills specific to a single test format or knowledge sampled by any given test. The set of skills encompassed in PIRATES therefore meets the definition of a strategy as defined by Alley and Deshler (1979). Teaching problem

solving strategies such as those described by Alley and Deshler and their colleagues at the KU-IRLD is not unique to the Kansas Institute. Authors in the field of educational psychology have also emphasized the value of strategy instruction (Brown, Campione, & Day, 1981; Flavell, 1977). Brown et al. point out that while average or above average achieving students may spontaneously create and use strategies when approaching academic tasks, below average students often do not. Direct instruction in strategy use may be necessary for these students if they are to learn and apply them.

Armbruster, Echols, and Brown (1983) distinguish techniques from strategies, stating that techniques become strategies when a student knows not only how to use a procedure but also when and where to apply it. Dansereau (1978) supports strategy instruction, suggesting that teaching methods which are situation-specific serve to limit the student's ability to transfer knowledge to other situations where it may be applicable. A learning strategies model is an approach which not only teaches acquisition of a strategy but includes activities for promoting generalization. Inclusion of generalization procedures when teaching a skill or strategy is advocated by writers such as Baer (1979) and Kuhn (1974) who state that generalization does not often occur unless it is taught. Results of generalization research reported by Schumaker et al. (1983) indicate specific generalization procedures must be used in order for a strategy learned in a resource classroom to be applied in the regular classroom. Researchers at the KU-IRLD have developed an instructional sequence designed to directly teach

a strategy as well as promote generalization and maintenance of the strategy in the regular, content classroom.

Instructional Procedures for Promoting Strategy Acquisition and Generalization

Various authors have emphasized the need for systematic and controlled procedures for instructing students in test-taking skills (Bangert-Drown et al., 1983; Brown, 1982; Fueyo, 1977; Markel, 1981). The central idea promoted by these authors is that students can learn the skills through a comprehensive, task-specific approach which includes a generous amount of teacher controlled and monitored practice. Markel (1981) promotes instruction of LD adolescents in test-taking which encompasses modeling, student self-vocalization, and practice. Relatedly, Stevens and Rosenshine (1981) state that handicapped pupils need instruction that carefully integrates characteristics of effective learning principles such as those listed above.

Deshler, Alley, Warner, and Schumaker (1981) have outlined special procedures that incorporate sound teaching techniques which are effective in promoting acquisition and generalization of learning strategies in LD adolescents. These procedures, or steps, are as follows:

1. Analysis of current learning habit. The purpose of this step is to assess how students currently solve or complete a particular task, and make them aware of any inefficiencies.

2. Describe the new strategy. The teacher describes the steps of the strategy being introduced and provides a rationale for its use.

3. Model the new strategy. The teacher models the strategy for the student while verbally describing each step.

4. Verbal rehearsal of strategy steps. In order for the student to learn the steps of the strategy at an automatic level, the student verbally rehearses each step. Verbal rehearsal is a self-instructional training procedure which aids learning cognitive strategies.

5. Student practices in controlled materials. The learning strategy is practiced with materials that are at the student's reading level. This allows the student to practice the strategy unhindered by content or reading material that is too difficult. The student practices the skill frequently until it is mastered.

6. Student practices in classroom materials. Once the strategy can be performed proficiently in controlled materials, the student is required to use it with grade level materials such as those used in the regular classroom. The student practices with these materials until proficiency is reached.

7. Generalization. Several techniques designed to promote generalization of the strategy are introduced after the student has mastered the above step. First, teach students to cue others such as the regular class teacher for reinforcement for using the strategy. Secondly, provide the student with diverse examples and opportunities to use the strategy. Thirdly, train loosely by varying formats and instructions given to students in order to approximate the variability of real learning situations. Other additional features of the generalization process include delaying reinforcement and simply telling the student to generalize.

The above instructional procedures will be used in this study to train subjects in the use of the PIRATES strategy. A detailed description of how these steps will be implemented in this study is found in Appendix A.

Why Should Mildly Handicapped (LD and EH) Students Be Taught a Test-Taking Strategy?

Most mildly handicapped adolescents spend a majority of the school day in mainstream classrooms (Friend & McNutt, 1984) where success is measured in most part by performance on teacher-made tests. Markel (1981) stresses the importance of teaching test-taking skills to adolescents receiving special education in order to assist them in succeeding academically in mainstream classes, and evidence does exist that many mildly handicapped students lack test-wiseness.

Alley et al. (1979) report that poor test-taking ability is one of the more commonly referred to characteristics of LD adolescents. Other writers (Havertape, 1976; Keogh, 1971) have described behaviors of LD students which may account for LD adolescents being perceived as poor test-takers. Keogh (1971) observed that these students often choose the first answer option in a multiple-choice question and do not monitor errors made when taking tests. Havertape (1976) noted LD students frequently do not pay attention to directions or systematically respond in problem solving situations. Given the above student characteristics, the frequency in which classroom tests are administered and the support of the efficacy of teaching test-taking skills, it can be assumed that LD secondary students may benefit from being taught a test-taking strategy.

Mainstreamed EH adolescents take classroom tests at the same frequency of other handicapped or nonhandicapped secondary students. However, unlike their LD counterparts, little is known about specific test-taking behaviors of EH adolescents in mainstream classes. It may be assumed that general school related characteristics such as impulsivity and poor independent work habits (Kauffman, 1981; Swift & Spivak, 1969) may hinder EH secondary students when taking classroom tests. The fact that most EH students are experiencing learning difficulties (Kauffman, 1981) may also indicate difficulty in expressing knowledge in testing situations.

Unfortunately, the lack of systematic, empirical research on the characteristics of, and effective intervention with, EH adolescents may provide the strongest rationale for including them in this study. Since so little published research exists (Hughes, Ruhl, & Korinek, 1985), investigating the efficacy of a test-taking strategy with EH secondary students may add to the limited body of knowledge about this population of handicapped students. The need for research with EH adolescents was recently highlighted by survey results reported by Epstein and Cullinan (1984). They surveyed experts in the field of emotional handicaps as to what aspects of this population were in critical need of further research. Programming for EH adolescents was ranked third of all research priorities.

Summary

This review of literature related to test-taking has been organized to address five questions. For purposes of continuity, these questions are used to provide a format for the following chapter summary.

What Is Test-Wiseness?

This term encompasses a variety of skills or cognitive abilities which allow the test-taker to utilize test format and characteristics, independent of knowledge and content, in order to score higher on a test. If, as writers in this area state, test performance is affected by TW, then students lacking these skills should be taught them.

Are Test Scores of Students Taught Test-Taking Skills Improved?

A review of data-based articles examining the effect of test-taking instruction on test scores yielded six studies meeting the selection criteria. Three studies reported significant results in favor of the experimental groups and two other investigations showed nonsignificant but positive results for the experimental groups. Nonsignificant results were reported in one study, with each group performing approximately the same.

While there appears to be some support for the effectiveness of teaching test-taking skills to regular, elementary students, similar evidence for handicapped secondary students is limited. Only one of the reviewed studies used this population.

What Skills Should Be Included When Teaching Students Test-Taking Skills?

Reviews of experimental and nonexperimental articles yielded a variety of test-taking skills that were used or suggested. Skills that consistently appear are listed in this chapter and included in the PIRATES unit which will be taught to subjects participating in the present study (see Table 3).

What Method Should Be Used To Teach Students Test-Taking Skills?

The proposed methodology for teaching test-taking unit PIRATES is a learning strategies approach. A strategies approach is designed to assist the student in acquiring behaviors as well as generalizing them to various settings (e.g., different secondary content courses) and formats. A specific instructional sequence designed by the KU-IRLD to teach adolescents learning strategies will be employed in teaching PIRATES to the subjects in this study.

Why Should Mildly Handicapped (LD and EH) Students Be Taught a Test-Taking Strategy?

Since most LD and EH students are mainstreamed they take teacher-made tests on the content covered in mainstream classes. Both LD and EH students exhibit behaviors which may affect their ability to take tests in a manner which reflects their full knowledge of the information assessed by the test. These students may therefore benefit from being taught a test-taking strategy which could positively affect academic success in the secondary school setting.

CHAPTER III
METHOD

The research method outlined in this chapter consists of applied behavior analysis procedures used in examining performance of a test-taking strategy by individual learners. The primary questions under investigation are

1. Do LD and EH secondary students acquire a test-taking strategy at a 90 percent mastery level when taught in a resource class setting?

2. After acquiring a test-taking strategy at 90 percent mastery level, are regular classroom test scores of LD and EH secondary students improved?

The methodology of this study is discussed according to the topics subject selection, subject description, setting, description of the test-taking strategy, instructional procedures, experimental design, reliability and validity procedures, and data analysis.

Subject Selection

Twelve subjects in the seventh and eighth grades were selected. Six of these 12 were identified as meeting the criteria for LD outlined by the State of Florida, and six were identified as meeting Florida criteria for EH. Additional criteria for inclusion in this study were as follows:

1. Subjects were receiving special education services on a part-time basis in a varying exceptionalities (V.E.) resource classroom.

2. Subjects were enrolled in a mainstream social studies or science course.

3. Subjects achieved a reading level of fourth grade or above on Metropolitan Achievement Test within the past year.

4. Subjects scored an 85 (full scale) or above on the most recent standardized intelligence test taken.

5. Subjects scored 50 percent or less on a pretest sampling aspects of test-wiseness that are included in the test-taking strategy to be taught.

Criteria 3 and 4 are suggested by researchers at the Kansas University Institute for Research in Learning Disabilities (Schumaker, Deshler, Alley, & Warner, 1983) as minimum ability and performance levels of students who could benefit from strategy training.

In order to obtain the 12 subjects for this study, two VE teachers were asked to make a list of their students whom they believed would meet the above criteria. Letters of parental consent for participation in the study were sent to students on the list. Once the letters were signed and returned, the cumulative records of the students were read to ascertain whether these students met the criteria for subject selection. Only 10 students met the I.Q. criteria, so the intelligence criteria was lowered 10 points to a full-scale score of 75. Fifteen students met the new criteria and were administered a pretest sampling

test-wiseness. All 15 scored below 50 percent, therefore meeting criteria 5. Twelve of the 15 students meeting all criteria were randomly selected for participation in the study.

Subject Description

Demographic information on subjects is displayed in an adaptation of the tabular form (see Table 4) recommended by the Council for Learning Disabilities Research Committee (Smith, Deshler, Hallahan, Lovitt, Robinson, Voress, & Ysseldyke, 1984) and follows subject description guidelines developed by professionals in the areas of learning disabilities and emotional handicaps (Lakin, 1983; Smith et al., 1984). Twelve middle school students, six labelled LD and six labeled EH, participated in the study. Nine of the subjects were in the eighth grade and three were in the seventh grade. Their ages ranged from 13 years, 3 months to 17 years, 2 months ($\overline{X}$ = 14.6, Median = 14.5). The participating students had I.Q. scores ranging from 78 to 118 ($\overline{X}$ = 92.5, Median = 91), and grade level reading scores ranging from 4.0 to 7.0 ($\overline{X}$ = 6.2, Median = 5.6). Subjects included two white females, seven black males, and three white males.

Setting

The research was conducted at Lincoln Middle School in the Alachua County Schools. Lincoln Middle School serves approximately 700 students. Each experimental session consisted of individual or small group instruction (one to four students) in one part of a varying exceptionalities(V.E.)classroom concurrent with instruction to other

Table 4

Subject Description

Subject	Gender	Age	Grade	Special Education Category	Ethnicity	I.Q.[a] (F-S)	Reading[b] Achievement
1	M	15.9	8	LD	W	91	5.6
2	F	13.1	7	LD	W	101	7.0
3	M	17.2	8	LD	B	80	4.5
4	M	14.5	8	LD	W	95	5.6
5	M	16.3	8	LD	B	84	4.0
6	M	14.9	8	LD	B	81	4.4
7	M	13.3	8	EH	B	118	5.0
8	F	14.5	8	EH	W	101	6.0
9	M	13.4	7	EH	B	101	5.7
10	M	16.5	8	EH	B	78	6.9
11	M	14.3	8	EH	B	90	4.0
12	M	13.8	7	EH	W	91	6.1

[a]Test(s) used: Weschler Intelligence Scale for Children--Revised (Administered within last three years).

[b]Test(s) used: Metropolitan Achievement Tests (Administered 4/84).

students not involved in the experiment by the V.E. teacher. Approximately 8 to 12 students are seen by the V.E. teachers per period. Neither V.E. teacher had an aide, making the adult-student ratio approximately 10:1. The classrooms themselves were of a standard size and contained both individual student desks and tables.

Description of the Test-Taking Strategy

The independent variable used in this study was the test-taking strategy PIRATES (see Table 5). This strategy was designed by task-analysis of test-taking behavior, logical sequencing of the tasks, and identifying the most frequently used test-wiseness (TW) principles reported in the literature. The mnemonic PIRATES was developed to increase the likelihood that subjects would remember the strategy components in an actual testing situation (Pressley, Levin, & Delaney, 1982). The following description includes the word that each letter of the acronym is paired with, the subject behaviors included within each component, and how each behavior will be measured. Additionally, examples of items used to measure specific TW principles are included when appropriate.

Table 5

The PIRATES Strategy

Behavioral Sequence	Measurement
NAME OF COMPONENT: P - Prepare	
1. Write PIRATES at the top of the test form.	1. Probes will be examined to see if the subject wrote PIRATES at the top of the test form.
2. Carefully read and follow the instructions (if any) at the beginning of the test.	2. Probes will be examined for written evidence that directions at the top of the probe were followed correctly. Example: "Write your name and the date on the top right hand corner of the test form. Indicate the class period in the top left hand corner."
3. Begin the test immediately remembering to work quickly, not taking too long on any one problem.	3. Students will be observed as to whether they began working within 90 seconds of being given a probe.
NAME OF COMPONENT: I - Instructions	
4. Read instructions for the first section carefully while underlining "key" words or phrases (see Appendix C). Make sure you understand what the instructions are telling you to do.	4. The probe will be examined for written evidence that the section instructions were followed (e.g., answers were marked in the manner designated in the instructions). Example: "Write T in front of those statements that are completely true. Write F in front of those statements that are completely false." Key words will be imbedded in the directions. This behavior will be assessed by calculating the ratio of how many key words were contained in the directions and how many the subject underlined. Example: "Indicate the best response by circling the letter preceding it."

Table 5--Continued

NAME OF COMPONENT: R - Read	
5. Read the first question/ statement carefully and all the way through.	5. No measure.
6. Read all answer options in sequence and mark out any obviously wrong answers.	6. Some questions will have obviously wrong options that occur sequentially before an obviously correct answer. This behavior will be assessed by calculating the ratio of opportunities to correctly mark out wrong options to the number of correctly marked out wrong options. Example: "People who study the ocean are called: a. plumbers b. oceanographers c. cooks d. bus drivers."
NAME OF COMPONENT: A - Answer or Abandon	
7. Appropriately mark the answer if you are sure it is correct. Double check to see if answers are marked in the appropriate place/ manner.	7. Section directions will include information on how the subject is to mark answers. This behavior will be assessed by examining the probe and computing the percentage of appropriately marked options (e.g., circled, place letter of option in the blank provided). Example: see number 4.
8. Abandon or skip a question if you are not sure of the correct answer. Indicate the question was skipped by placing a small check (✓) to the left of the item skipped.	8. Some questions will not have a right answer because the question will be about a fictitious word or concept. Therefore, the student should not be sure of the correct answer and should check it in the appropriate manner. This behavior will be assessed by calculating the percentage of "no answer" questions whose number is circled. Example: "The green fredent a. never hunts at night b. always mates at three years c. usually burrows when frightened d. never eats its young."

Table 5--Continued

NAME OF COMPONENT: T - Turn Back	
9. Turn back to step 5 and go to the next question.	9. No measure.
NAME OF COMPONENT: E - Estimate	
10. When all sections of the test have been completed, return to the checked questions and estimate the answers using available cues (see Appendices B and D).	10. When returning to "abandoned" questions the student will estimate or guess the correct option using test-wiseness strategies that utilize cues that are dependent on the test constructor. These cue strategies are listed in Appendices B and D. Each cue will occur at least four times within a probe and will be imbedded in multiple choice questions or true-false statements which have no correct answer. This behavior will be assessed by calculating the percentage of how many times the options containing a cue was selected. Example: (length of option/most qualified option) Dinophos are dangerous because: a. people like them. b. they are hard to see. c. they kill certain insects which are helpful to many food crops. d. they live for a long time.
NAME OF COMPONENT: S - Survey	
11. When finished estimating, survey the test to make sure all questions have been answered.	11. Probes will be examined to assess whether all questions have been answered.

Table 5--Continued

12. Do not change answers unless the question/ answer was misread or very sure the first response is incorrect.	12. The experimenter will observe whether the subject changed any answers and the student will be asked to explain the reason for the change. Additionally, the probe will be examined for any evidence of erasures and the subject will be questioned as to the reasons for them.

Instructional Procedures

The test-taking strategy PIRATES was taught to the subjects participating in this study by instructional procedures designed for promoting skill acquisition and generalization in LD adolescents (Deshler, Alley, Warner, & Schumaker, 1981). The procedures developed by Deshler et al. were modified for the purposes of this study. The following section lists seven instructional steps and identifies the main purpose of each (see Appendix A for a more detailed description).

Step 1: Test

Purpose: To provide a baseline of each subject's use of the test-taking skills contained in each component of the PIRATES strategy.

Step 2: Describe

Purpose: To describe the behaviors included in the PIRATES strategy and to provide a rationale for using the strategy.

Step 3: Model

Purpose: To demonstrate, while verbally describing, the PIRATES strategy in its entirety.

Step 4: Verbal Rehearsal

Purpose: To ensure subjects can name the components of the PIRATES strategy.

Step 5: Practice and feedback in controlled materials

Purpose: To provide practice in using the test-taking behaviors included in the PIRATES strategy. For instructional purposes components were grouped and practiced in two clusters as follows: Cluster 1, behaviors 1 through 9; Cluster 2, behaviors 10 through 12.

Step 6: Practice and feedback in content materials

Purpose: To provide practice using all the test-taking behaviors included in the PIRATES strategy. This step will also be used to assess whether subjects are proficient in the use of the strategy.

Step 7: Promote Generalization

Purpose: To promote generalization of the PIRATES strategy to regular classroom testing situations.

Experimental Design

A single subject research design was used and is considered appropriate for this proposed study because (a) the use and development of the strategy used in this study is at the exploratory stage (Mercer, Bott, & O'Shea, 1984), and (b) individual or small group instruction is the most frequent instructional arrangement used in a special education setting (Lue, 1981; Schumaker & Clark, 1982). As Strain, McConnell, and Codisco (1983) state, there is a "conceptual compatibility between special education practices and single-subject design requirements. For example, in both endeavors we typically are

concerned with assessment of individuals, careful specification of an intervention plan, and evaluation based upon the progress a specific child makes" (pp. 401-41). Finally, the use of repeated measures associated with a single subject research approach is a more reliable way to observe subject performance than the use of a single measure which is commonly used in group designs (Haring, Lovitt, Eaton, & Hansen, 1978).

The type of single subject method used in this study is a multiple baseline design across subjects. Four groups of three subjects each were used. Two of the groups were composed of three LD students each and the other two groups were composed of three EH students each.

A multiple-baseline design was selected because (a) replication of effects across several subjects will enhance confirmation and generality of results, and (b) the use of reversal designs (e.g., A-B-A) was not appropriate as the behavior to be learned is not readily reversible (Alberto & Troutman, 1982; Tawney & Gast, 1984). Additionally, multiple-baseline designs are capable of providing evidence of a functional relationship between treatment and changes in behavior. Within the context of a multiple-baseline design this relationship is assumed if "each dependent variable in succession shows a change when and only when the independent variable is introduced" (Alberto & Troutman, 1982, p. 148). Examples of the research design are found in Figures 1 and 2 and illustrate condition sequences for the two sets of data used in this study: acquisition and generalization.

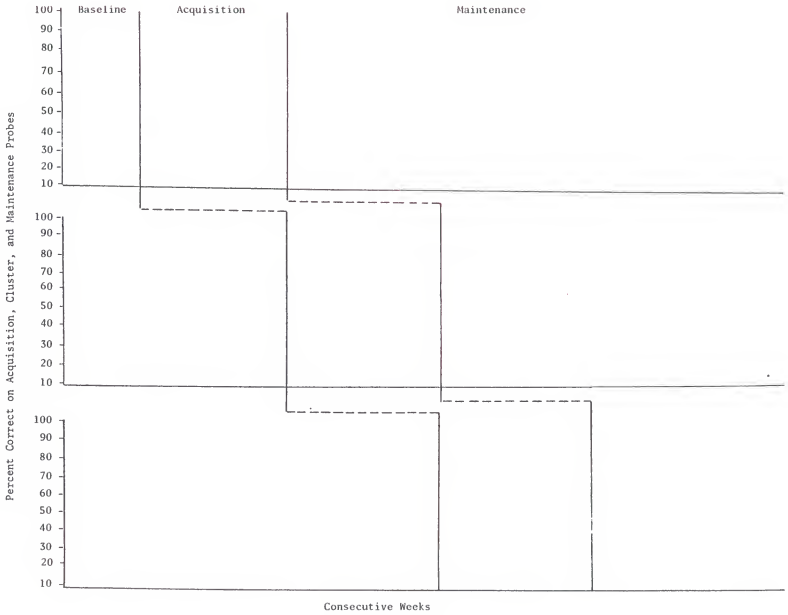

Figure 1. Format for Recording Acquisition Data.

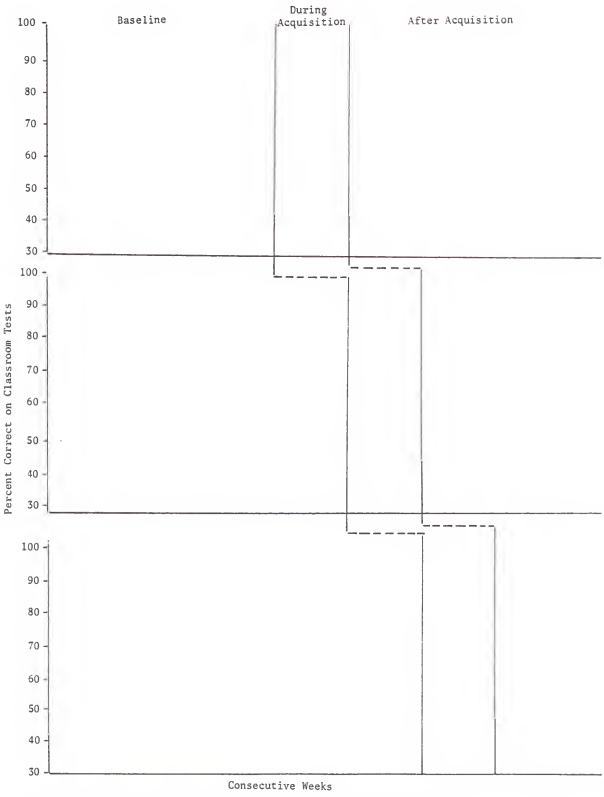

Figure 2. Format for Recording Generalization Data.

Acquisition Data

Acquisition data reflect subject performance of components of the PIRATES strategy as measured by experimenter made probes (see Figure 1). Acquisition data are used to answer the first experimental question of whether subjects can learn or acquire PIRATES to a proficient level. The acquisition data are displayed in three experimental conditions: baseline, acquisition, and maintenance.

Baseline condition

Two experimenter-developed pretests or probes sampling all test-taking behaviors included in the strategy PIRATES were administered in a two-week period to all subjects. Four subjects (2 EH and 2 LD) exhibiting stable scores on the initial measures started the acquisition condition. When these first four subjects reached proficiency (90 percent mastery) in all components of PIRATES, a third pretest probe was administered to the remaining subjects. Four subjects who exhibited stable scores on the three pretest probes began the acquisition condition. When this second set of subjects reached proficiency in acquiring the PIRATES strategy, the remaining four subjects were administered a fourth pretest probe subsequent to starting the acquisition condition.

Acquisition condition

During this condition subjects were taught the PIRATES strategy using the instructional procedures in Appendix A. The strategy was taught in two clusters, with subjects required to reach mastery in cluster one behaviors before receiving instruction in cluster two behaviors. Cluster one probes sampled behaviors one through nine of the PIRATES strategy, and cluster two probes emphasized behaviors 10 through 12. Equivalent forms of cluster one and two probes were

administered, interspersed with instruction, until mastery (90 percent) was reached.

After mastery was demonstrated on cluster two skills, equivalent forms of the acquisition probe, which sampled all behaviors included in the PIRATES strategy were administered. These probes are equivalent to the pretest probes administered during the baseline condition.

Maintenance condition

After subjects acquired mastery of the PIRATES strategy, maintenance probes were administered every several weeks in order to assess whether subjects still exhibited behaviors taught during the acquisition condition. These probes are equivalent to both pretest and acquisition probes.

Generalization Data

Generalization data consist of percent correct on unit tests administered in a mainstream classroom (see Figure 2). These data were used to address the second experimental question of whether subjects who have acquired the PIRATES strategy improve their performance on mainstream classroom tests. The generalization data are displayed in three experimental conditions: baseline, during acquisition, and after acquisition.

Baseline condition

These data are scores on mainstream classroom tests taken before training in PIRATES was introduced. A minimum of five data points is recorded in this condition.

Acquisition condition

These data are scores on mainstream classroom tests taken during the time that the strategy was taught. A minimum of two data points is recorded during this condition.

After acquisition condition

These data are scores on mainstream classroom tests taken after subjects acquired the PIRATES strategy. The only procedures subjects were exposed to during this condition were maintenance probes and one generalization session (see Appendix A).

Reliability and Validity Procedures

Several procedures were used during the experiment in order to establish reliability and validity of measurement and procedures. These were content validity and equivalency of probes used to measure use of the PIRATES strategy, interscorer agreement, and procedural reliability checks.

Content Validity and Equivalency

Before baseline measures of subject use of test-wiseness skills included in the PIRATES strategy were taken, the probes developed by the experimenter to measure these skills were examined by three professionals in special education and/or tests and measurement for content validity. All three professionals were asked to examine the instruments used in the study and answer the following questions:

1. Do the probes assess the behaviors they are said to by the experimenter?
2. Are the probes equivalent in the number and types of behaviors they assess?
3. Are the formats of the probes equivalent?

The individuals assessing content validity and equivalency of the probes were provided with test forms, a list of which behavior was being sampled by each item, and the scoring procedures. All three professionals agreed that the probes measured the behaviors they purported to and that the forms of the probes were equivalent in the types of behaviors sampled and in format.

Interscorer Agreement

During the experiment a professional in the field of special education randomly selected nine acquisition probes previously scored by the experimenter. In order to establish interscorer agreement, this individual was provided with scoring procedures used by the experimenter (see Appendix E) and asked to independently score each of the nine probes. Percent of agreement was then calculated by dividing the number of possible agreements for the probes into the actual number of agreements. This calculation yielded an agreement score of 96 percent which is considered sufficient to establish the reliability of scoring procedures used in this experiment.

Procedural Reliability

In order to ensure procedural reliability an independent observer randomly observed instructional sessions to ascertain whether instructional procedures (see Appendix A) were consistent between sets of subjects and whether probes administered in both baseline and acquisition conditions were completed by subjects without experimenter assistance. Observations by the independent observer confirmed that procedures were consistent and that subjects completed probes independently.

Data Analysis

Question 1

Do LD and EH secondary students acquire a test-taking strategy at a 90 percent mastery level when taught in a resource setting?

Data used to answer question 1

Data points representing percentage correct on pretest, cluster, and acquisition probes were charted (see Figure 1).

Analysis of data

Visual inspection of data displayed on charts similar to Figure 1 was used to identify performance trends of individual subjects' acquisition of the PIRATES strategy. Additionally, data obtained were analyzed by assessing whether the set criteria for mastery (90 percent) on the acquisition probes were reached.

Question 2

After acquiring a test-taking strategy at 90 percent mastery level, are regular classroom test scores of LD and EH secondary students improved?

Data used to answer question 2

Percentage correct on regular classroom tests was the type of data used to answer question 2.

Analysis of data

Data were displayed on charts similar to Figure 2 and were analyzed by two methods: comparison of between condition trends and between condition levels (Tawney & Gast, 1984). Trends for each condition were calculated using a split-middle method (White & Haring, 1980) which provides a reliable estimate of trend when analyzing variable data

patterns. Level change between conditions was calculated using both means and medians of the two adjacent conditions in order to find the most conservative measure of change. Additionally, significance of any gains on test scores by individual subjects was analyzed as to the functional or practical impact of the gain (e.g., the effect of the gain on letter grades).

CHAPTER IV
RESULTS

In Chapter IV results are reported in response to the experimental questions posed in this study. This chapter includes a summary of the overall results for the first experimental question, each subject's individual data summaries for the second experimental question, and a summary of results for the second experimental question.

Summary of Results for Question One

The first question was one of acquisition and asked whether subjects in this study acquire the PIRATES strategy at a 90 percent mastery level. Figures 3 through 6 present results of the multiple-baseline across subjects design used to answer this question. The data displayed on these figures represent percentage correct scores on the probes (i.e., baseline, cluster, acquisition, and maintenance probes) used to measure acquisition of the strategy (raw data are displayed in Appendix F). Due to the relative uniformity of acquisition rates across subjects, a summary of the combined results is presented here.

During baseline condition, no subject scored above 34 percent correct on baseline probes (range 18-34 percent). During the acquisition condition, in which the strategy PIRATES was taught, all but one subject (Subject 6) scored 90 percent or above on at least one acquisition probe.

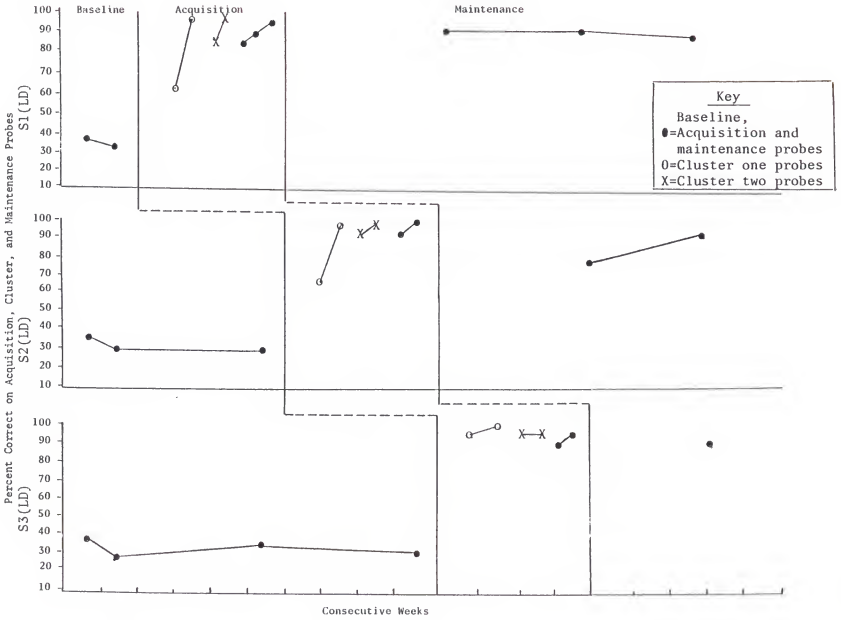

Figure 3. Acquisition Data for Subjects 1, 2, and 3.

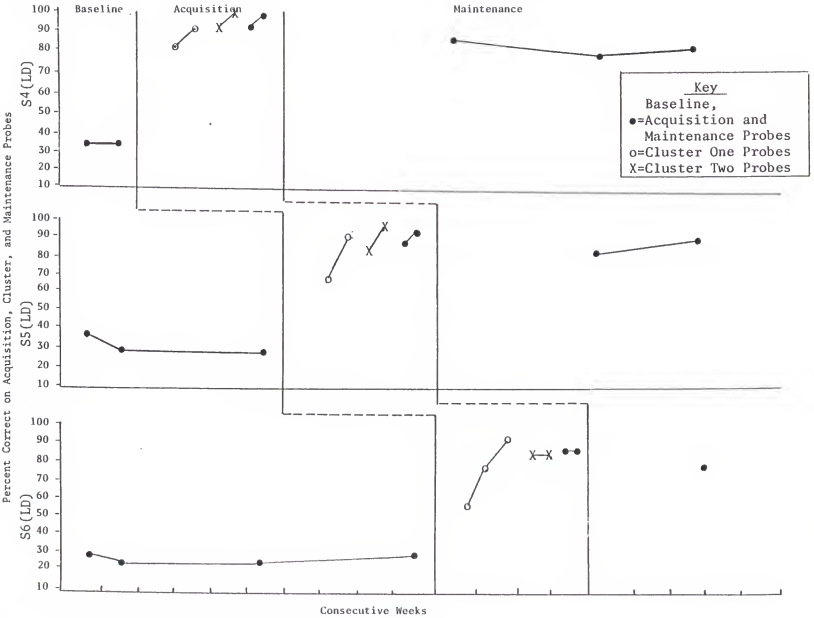

Figure 4. Acquisition Data for Subjects 4, 5, and 6.

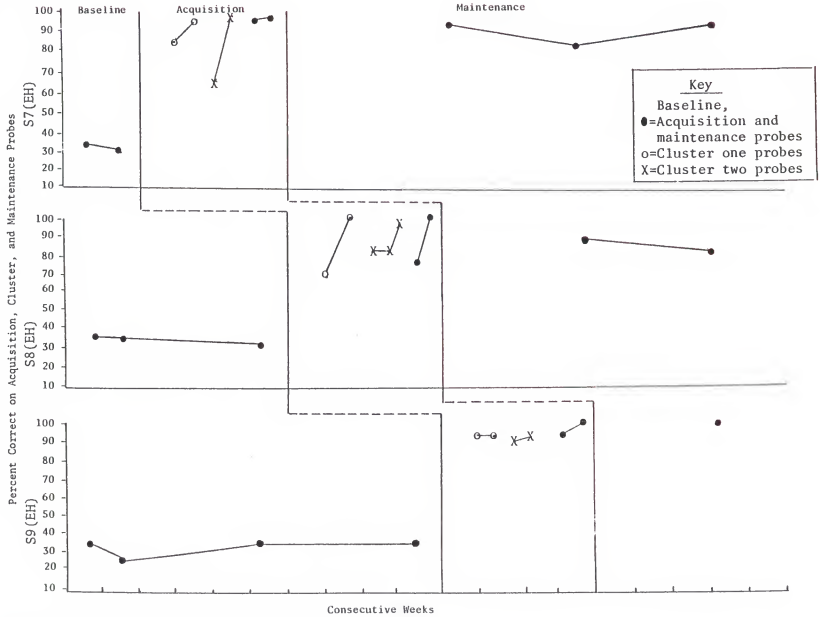

Figure 5. Acquisition Data for Subjects 7, 8, and 9.

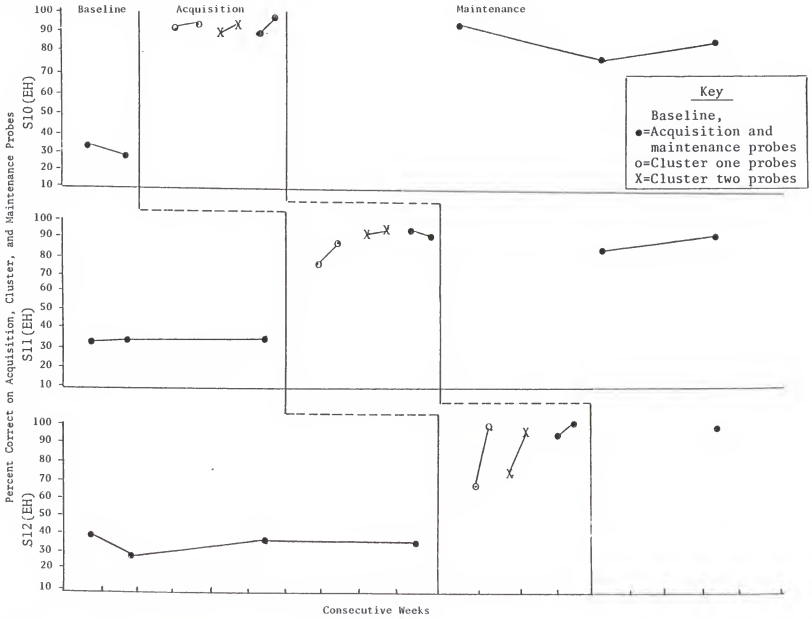

Figure 6. Acquisition Data for Subjects 10, 11, and 12.

Post acquisition maintenance probes measuring all test-taking behaviors sampled by the PIRATES strategy were administered at four-week intervals to determine if acquisition was stable over time. The staggering of intervention dictated by a multiple-baseline design and an externally imposed time constraint (i.e., the end of school) resulted in the first set of subjects completing three maintenance probes, the second set completing two, and the third set of subjects completing one probe. Results displayed in Figures 3 through 6 show that while scores on maintenance probes were slightly lower over time, most were still performing only a little below or at mastery. Eight subjects scored 86 percent or above on the final maintenance probe they were administered. The remaining four subjects, 4, 6, 7, and 10, scored 78, 82, 84, and 84 percent, respectively.

The above results indicate that most subjects in this study did acquire, at 90 percent mastery level, the PIRATES strategy. Of the 12 subjects, 11 scored 90 percent or above on at least one acquisition probe. The one subject (Subject 6) scoring lower than 90 percent had frequently been absent during the four week instruction period and missed several practice and instructional sessions. Additionally, it appears that acquisition is maintained over time albeit at a slightly lower level.

A functional relationship between treatment and changes in behavior (i.e., use of the PIRATES strategy) was established. Within the context of the multiple-baseline design used, this relationship was established when an obvious positive change in the dependent

variable (i.e., scores on probes) occurred with the introduction of the independent variable (i.e., instruction on PIRATES). All subjects showed a stable baseline followed by an easily discernable increase in scores after instruction.

Individual Data Summaries for Question Two

Results of the multiple-baseline design used to address the second experimental question are displayed in Figures 7 through 10. This question was one of generality and investigated whether subjects who have acquired the test-taking strategy PIRATES improve test scores in a mainstream classroom (i.e., social studies or science). The format for this discussion is by individual subject and is presented as follows: subject description, comparison of mean and median change between conditions (see Table 6) and analysis of trends between conditions (Figures 7-10). It should be noted that the during acquisition and after acquisition conditions were combined when calculating the mean and median. This was done for two reasons. First, there were only one or two data points for the during acquisition condition. Secondly, subjects had acquired, to some extent, all or part of the strategy when they took a classroom test during this condition. This was especially true for tests taken after the second week of instruction, as most subjects had acquired, at a mastery level, the first cluster of test-taking skills included in PIRATES.

Subject 1 Data Summary

Subject 1 was a 15 year old, eighth grade, white male labeled learning disabled. The subject was reported to have an I.Q. of 91, and a reading level of 5.6.

Table 6

Differences in Mean and Median Scores Between Conditions

Subject	Mean		Median		Difference	
	Before	During/After	Before	During/After	Mean	Median
S1	65	72	64	74	+ 7	+10
S2	57	68	54	64	+11	+10
S3	54	65	58	65	+11	+ 7
S4	60	75	59	80	+15	+21
S5	55	72	56	78	+17	+22
S6	50	74	52	80	+24	+28
S7	54	61	50	64	+ 7	+14
S8	55	65	62	69	+10	+ 7
S9	55	81	61	81	+26	+20
S10	58	59	58	58	+ 1	0
S11	58	68	60	64	+10	+ 4
S12	62	74	60	73	+12	+13

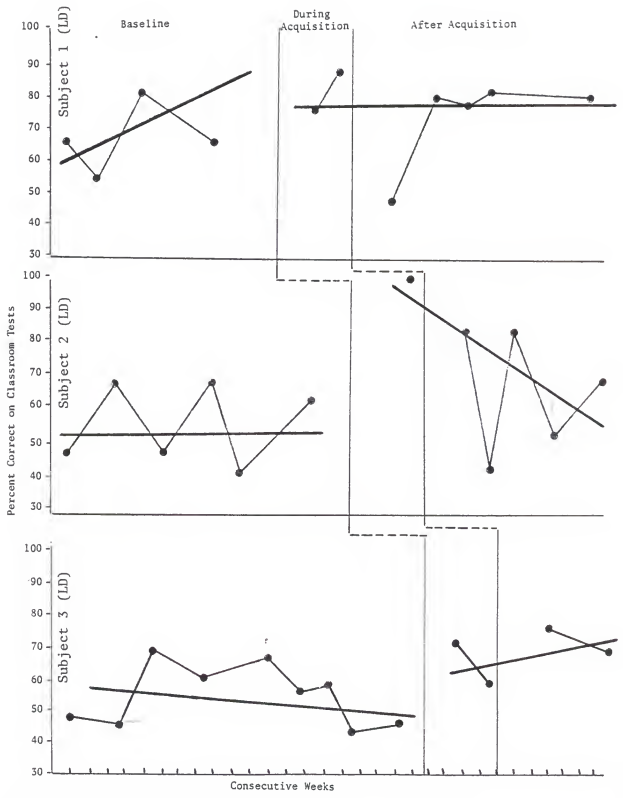

Figure 7. Generalization Data for Subjects 1, 2, and 3.

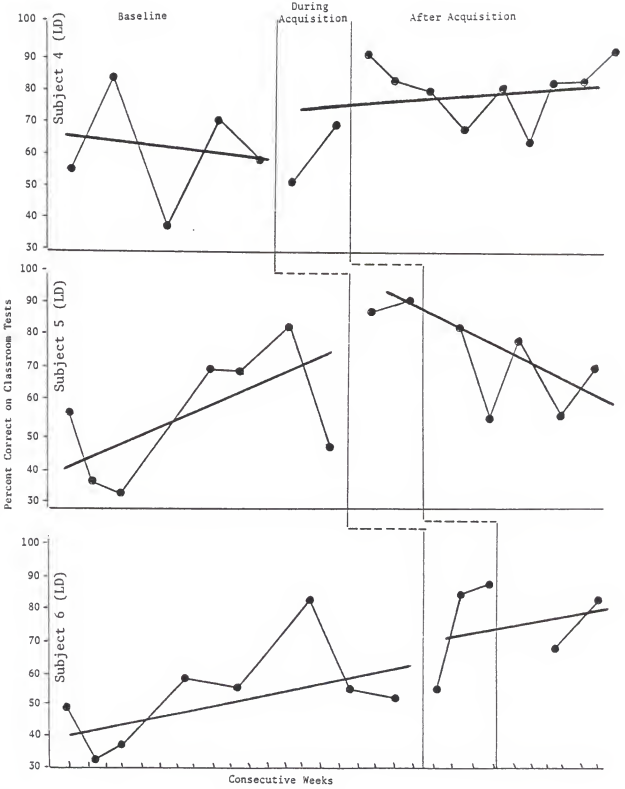

Figure 8. Generalization Data for Subjects 4, 5, and 6.

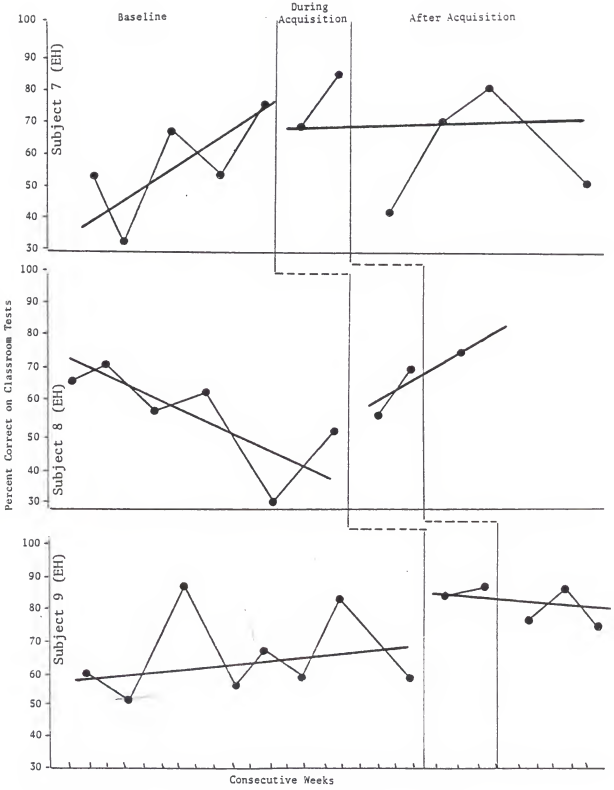

Figure 9. Generalization Data for Subjects 7, 8, and 9.

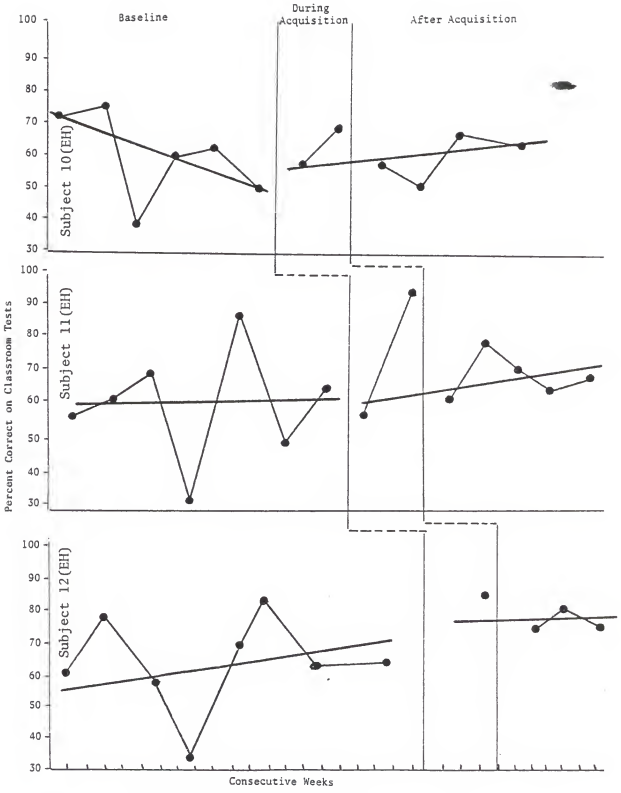

Figure 10. Generalization Data for Subjects 10, 11, and 12.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a mainstream science class by Subject 1 is presented in Figure 7. The mean and median scores for the baseline condition were 65 and 64 percent, respectively. Mean and median scores calculated to include both during and after acquisition conditions were 72 and 74, respectively. The mean and median differences between conditions shows an increase of 7 and 10 percentage points, respectively (see Table 6).

Analysis of trends between conditions

Visual analysis of trend lines plotted by the split middle method (White & Haring, 1980) indicates an accelerating trend for the baseline condition, with scores ranging from 53 to 80 percent (see Figure 7). The trend line for the during/after condition shows a flat trend. The data during this condition ranged from 44 to 85 percent, however, if the single outlier is excluded, the data in the second condition are much less variable (70 to 85 percent) than that in the first condition.

Subject 2 Data Summary

Subject 2 was one of the two females in the study. She was 13 years old, in the seventh grade, and labeled learning disabled. She read at a seventh grade reading level and was reported to have an I.Q. of 101.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a social studies class by Subject 2 is found in Figure 7. For the baseline

condition, the mean score was 57 and the median score was 54. Mean and median scores for the during/after acquisition condition were 68 and 64 percent, respectively, showing a mean increase of 11 percent and a median increase of 10 percent between conditions (see Table 6).

Analysis of trends between conditions

Visual analysis of trend lines for Subject 2 indicates a flat trend for the baseline condition with scores ranging from 40 to 68 percent (see Figure 7). The trend line for the during/after condition also shows a decreasing trend but at a higher level. The data during this condition were more variable than baseline, with scores ranging from 50 to 100.

Subject 3 Data Summary

Subject 3 was a 17 year old, eighth grade, black male, labeled learning disabled. He was reported to have an I.Q. of 80 and a reading level of 4.5.

Comparison of mean and median change between conditions

Percentage correct on tests taken in a social studies class by Subject 3 are presented in Figure 7. Scores of 54 and 58 percent were the mean and median scores, respectively, for the baseline condition. Mean and median scores for the during/after acquisition condition were both 65 percent, showing a mean increase of 11 percent and a median increase of 7 percent between conditions (see Table 6).

Analysis of trends between conditions

A slightly decaying trend for the baseline condition with scores ranging from 43 to 70 percent is obvious from a visual analysis of

trend lines for Subject 3 (see Figure 7). The trend line for the during/after acquisition condition shows a slightly accelerating condition starting at a higher level than baseline. Data for this condition were less variable than for the baseline condition with scores ranging from 58 to 71 percent.

Subject 4 Data Summary

Subject 4 was a 14 year old, eighth grade, white male labeled learning disabled. He was reported to have an I.Q. of 95 and a reading level of 5.6.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a science class by Subject 4 is presented in Figure 8. The mean and median scores for the baseline condition were 60 and 59 percent, respectively. Mean and median scores for the during/after acquisition scores were 75 and 80 percent, respectively, showing a mean increase of 15 percent and a median increase of 21 percent between conditions (see Table 6).

Analysis of trends between conditions

Visual analysis of trend lines indicates a slightly decaying trend for the baseline condition with scores ranging from 36 to 82 percent (see Figure 8). The trend line for the during/after condition shows an accelerating trend with scores ranging from 50 to 96. Further visual analysis shows a more stable trend at a higher level after acquisition had occurred than during baseline.

Subject 5 Data Summary

A 16 year old, eighth grade, black male labeled learning disabled was Subject 5. He was reported to have an I.Q. of 84 and a reading level of 4.0.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a social studies class by Subject 5 is presented in Figure 8. Subject 5 obtained a mean of 55 and a median of 56 for scores during the baseline condition. His mean and median scores for the during/after acquisition condition were 72 and 78, respectively, showing a mean increase of 17 percent and a median increase of 22 percent between conditions (see Table 6).

Analysis of trends between conditions

An increasing trend during the baseline condition with scores ranging from 32 to 80 percent is apparent from visual analysis of trend lines for Subject 5 (see Figure 8). The trend line for the during/after condition shows a decaying trend after a large initial increase on the test scores. Data during this condition were less variable than for the baseline condition with scores ranging from 56 to 86 percent.

Subject 6 Data Summary

Subject 6 was a 14 year old, eighth grade, black male labeled learning disabled. The subject was reported to have an I.Q. of 81 and a reading level of 4.4.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a social studies class by Subject 6 is presented in Figure 8. The mean and median scores for the baseline condition were 50 and 52 percent, respectively. Mean and median scores for the during/after acquisition condition were 74 and 80 percent, respectively, showing a mean increase of 24 percent and a median increase of 28 percent between conditions (see Table 6).

Analysis of trends between conditions

Visual analysis of trend lines for this subject shows a slightly increasing trend for the baseline condition with scores ranging from 30 to 80 percent (see Figure 8). The trend line for the during/after acquisition condition shows an almost identical acceleration of trend but at a slightly higher level, with scores ranging from 58 to 84 percent. If the data point recorded during the first week of instruction was discounted in consideration of inadequate time to show effect, the trend would be flat but at a much higher level. This would indicate a more dramatic effect than the existing comparison of trend.

Subject 7 Data Summary

Subject 7 was a 13 year old, eighth grade, black male labeled emotionally handicapped. He was reported to have an I.Q. of 118 and a reading level of 5.0.

Comparison of mean and median change between conditions

Percentage correct on tests taken in a science class by Subject 7 are presented in Figure 9. Scores of 54 and 50 percent were the mean and median scores, respectively, for the baseline condition.

Subject 7's mean and median scores for the during/after acquisition condition were 61 and 64 percent showing a mean increase of 7 percent and a median increase of 14 percent between conditions (see Table 6).

Analysis of trends between conditions

Resulting from visual analysis, Subject 7's trend line shows an increasing trend during the baseline, with scores ranging from 30 to 72 percent (see Figure 9). The trend line for the during/after acquisition condition indicates an almost flat trend starting at about the same level as the last data point in the baseline condition. Data in the during/after acquisition condition were almost as variable as in baseline with scores ranging from 45 to 80 percent.

Subject 8 Data Summary

Subject 8 was a 14 year old, white female, in the eighth grade. She was labeled emotionally handicapped and reportedly had an I.Q. of 101 and a reading level of 6.0.

Comparison of mean and median change between conditions

Percentage correct on tests taken in a social studies class by Subject 8 is presented in Figure 9. For the baseline condition, her mean score was 55 and median score was 62. Mean and median scores for the during/after condition were 65 and 69, respectively, showing a mean increase of 10 percent and a median increase of 7 percent between conditions (see Table 6).

Subject 8 missed several opportunities to take tests during the second condition. One test was missed due to an absence and subsequent failure to take a make-up exam. Two tests were missed because Subject 8 was suspended for the last few weeks of school. The sparcity of

data in the second condition inhibits a comparison of mean and median scores between conditions.

Analysis of trends between conditions

Visual analysis of trend lines for Subject 8 shows a decreasing trend during baseline with scores ranging from 30 to 70 percent (see Figure 9). The trend line for the during/after acquisition condition shows an increasing trend. Data recorded during this condition shows decreased variability of scores (55 to 72 percent). While a clear difference in data trend exists between conditions, the relatively few scores included in the second condition warrant caution in the interpretation of these data.

Subject 9 Data Summary

Subject 9 was a 13 year old, seventh grade, black male labeled emotionally handicapped. The subject was reported to have an I.Q. of 101 and a reading level of 5.7.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a social studies class by Subject 9 is presented in Figure 9. The mean and median scores for the baseline condition were 55 and 61 percent, respectively. Both the mean and median scores for the during/after acquisition condition were 81 percent, showing a mean increase of 26 percent and a median increase of 20 percent (see Table 6).

Analysis of trends between conditions

Through visual analysis of trend lines for Subject 9, a slightly increasing trend for the baseline condition with scores ranging from

54 to 89 percent (see Figure 9) is apparent. The trend line for the during/after condition shows a slightly decaying trend, however, the trend line begins at a higher level than in the baseline and does not fall below the baseline condition trend line. The data in the second condition is much less variable than in baseline with scores ranging only from 77 to 83 percent.

Subject 10 Data Summary

Subject 10 was a 16 year old, eighth grade, black male labeled emotionally handicapped. He had an I.Q. of 78 and a reading level of 6.9.

Comparison of mean and median change between conditions

Presentation of percentage correct on tests taken in a social studies class by Subject 10 is in Figure 10. The mean and median scores for the baseline condition for Subject 10 were both 58 percent. His mean and median scores for the during/after acquisition condition were 59 and 58 percent, respectively, showing a mean increase of 1 percent and no increase in medians between conditions (see Table 6).

Analysis of trends between conditions

Visual analysis of trend lines indicates a decreasing trend for the baseline condition with scores ranging from 36 to 72 percent (see Figure 10). The trend line for the during/after condition shows a slightly increasing trend which starts near the same level as the end of the baseline trend. Data in the second condition were less variable than during baseline with scores ranging from 48 to 68 percent.

Subject 11 Data Summary

Subject 11 was a 14 year old, eighth grade, black male. The subject, labeled emotionally handicapped, was reported to have a reading level of 4.0 and an I.Q. of 90.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a science class by Subject 11 is found in Figure 10. His mean and median scores for the baseline condition were 59 and 60 percent, respectively, as compared to 68 and 64 percent for the during/after condition. Comparison between conditions showed a mean increase of 10 percent and a median increase of 4 percent (see Table 6).

Analysis of trends between conditions

An almost flat trend during baseline with scores ranging from 31 to 85 percent results from visual analysis of Subject 11's trend lines (see Figure 10). The trend for the during/after acquisition condition shows a slightly greater acceleration with scores ranging from 55 to 92 percent.

Subject 12 Data Summary

Subject 12 was a 13 year old, seventh grade, white male labeled emotionally handicapped. He was reported to have an I.Q. of 91 and a reading level of 6.1.

Comparison of mean and median change between conditions

A display of percentage correct on tests taken in a science class by Subject 12 is presented in Figure 10. The mean and median scores for the baseline condition were 62 and 60 percent, respectively. Mean and median scores for the during/after acquisition condition were 74 and 73 percent, respectively, showing a mean increase of 12 percent and a median increase of 13 percent between conditions (see Table 6).

Analysis of trends between conditions

Visual inspection of Figure 10 indicates a slightly accelerating trend for the baseline condition with scores ranging from 32 to 80 percent. The trend line for the during/after acquisition condition shows a flat trend but at a higher level and with less variability (70 to 80 percent).

Summary of Results for Question Two

A summary of mean and median scores for the two conditions (baseline and during/after acquisition) is presented in Table 6. The mean and median difference between these two conditions is also displayed in Table 6. With the exception of Subject 10, all subjects showed an increase in both mean and median scores between conditions. Mean increases for subjects ranged from 1 percent to 26 percent and the range for median increase was 0 to 28 percent. The average increase between conditions was 12.5 percent and 13 percent for the mean and median, respectively.

A more functional way to view changes between conditions is to analyze differences in terms of letter grades. A summary of changes in obtained grades between conditions is found in Table 7. These grades are based on the current grading system used at the school attended by the subjects. When converting the means to letter grades and comparing the grades across the two conditions, eight subjects increased a letter grade, two subjects increased two letter grades, one subject increased three letter grades, and one subject remained the same. Using median scores converted to letter grades for comparison

Table 7

Differences in Letter Grades Between Conditions

Subject	Mean		Median	
	Before	During/After	Before	During/After
1	D	C	D	C
2	F	D	F	D
3	F	D	F	D
4	D	C	F	B
5	F	C	F	C
6	F	C	F	B
7	F	D	F	D
8	F	D	D	D
9	F	B	D	B
10	F	F	F	F
11	F	D	D	D
12	D	C	D	C

of performance across conditions, five subjects increased a letter grade, two subjects increased two letter grades, two subjects increased three letter grades, and three subjects remained the same. Using the most conservative difference, either mean or median, for individual subjects, six subjects increased a letter grade, three increased two grades, and three subjects remained the same.

Trend lines between conditions for each subject are displayed in Figures 7 through 10. These lines were used to assist in visual inspection of direction (e.g., scores accelerated after the independent variable was introduced) and immediate effects of intervention (e.g., did the trend line begin at a higher level than the end of the trend line from the previous condition). In the present study, five subjects (Subjects 3, 4, 8, 10, 11) showed flat or decreasing trends during baseline followed by an increasing trend line in the next condition. Subjects 2, 5, and 9 had decreasing trends in the during/after condition but the trend began at a higher level than the end of the trend for the previous condition. The remaining subjects (Subjects 1, 6, 7, 12) did not show improving trends or trends starting at higher levels for the during/after condition.

Summary

Results of this investigation showed subjects acquired the test-taking strategy PIRATES at a 90 percent mastery level, when taught using an instructional procedure designed to promote acquisition and generalization. All subjects but one acquired the strategy at the prescribed mastery level. Additionally, scores on post acquisition

probes administered at four week intervals indicate that acquisition is maintained over time at a slightly lower level.

In order to ascertain the effect of subject's acquisition of the PIRATES strategy on their performance on mainstream classroom tests, the mean and median scores on these tests were used to compare baseline performance with during/after performance. With the exception of one subject, the during/after condition for all subjects had a higher mean and median. In analyzing trend lines, eight subjects had either accelerating trends or trends indicating an obvious initial increase in test scores after the strategy was acquired. Finally, and perhaps of greater functional impact, when means and medians between conditions were transposed to letter grades and the most conservative of these measures used, results showed that 9 of 12 subjects increased at least one letter grade between conditions.

CHAPTER V
DISCUSSION

The purpose of the present study was to investigate the use of a test-taking strategy with emotionally handicapped (EH) and learning disabled (LD) adolescents. Two experimental questions were addressed. First, can EH and LD middle school students acquire skills in using a test-taking strategy? Secondly, do students who have acquired these skills improve their test scores in mainstream classrooms?

Using a set of instructional procedures developed through the Kansas University Institute for Research in Learning Disabilities (KU-IRLD) subjects were taught a test-taking strategy. The strategy, PIRATES, included skills generally used to promote test-wiseness. Subject performance on probes designed to measure acquisition of the strategy was used to answer the first experimental question that, overall, skills in using the strategy were acquired and maintained.

In order to answer the second experimental question, subjects' scores on tests administered in regular classrooms over a period of seven months were recorded. Baseline conditions lasted approximately three to five months depending upon which group the subject was assigned. After the skills had been acquired by all subjects, data (i.e., test scores) were analyzed to determine mean and median

differences between baseline and during/after acquisition conditions. Additionally, data trends between conditions were compared to ascertain differences, if any, in data direction and level. Results indicate that generally scores did improve after the strategy was acquired.

This chapter presents a discussion of the findings of this investigation. It is organized according to the two experimental questions posed in this study, limitations of the study, and recommendations for future research.

Question 1: Do LD and EH Secondary Students Acquire a Test-Taking Strategy at 90 Percent Mastery When Taught in a Resource Setting?

Schumaker et al. (1983), in summarizing research by the KU-IRLD, stated that the instructional procedures designed to promote learning strategy acquisition for LD adolescents (see Appendix A) have proven effective across a variety of strategies. Results of the present study add support to this contention. Taught through procedures similar to those from KU-IRLD all subjects, with the exception of one, acquired the PIRATES strategy at a 90 percent or above mastery level. The subject who did not acquire the strategy at the prescribed mastery level was absent more frequently than other subjects which may have hindered his progress. Interestingly, this subject showed comparatively high gains in actual test scores when mean and median scores across conditions were compared.

Results showing that use of the PIRATES strategy was maintained over time may be, in part, due to the instructional procedures used.

After administration and scoring of the first maintenance probe, each subject was asked which aspects of instruction had helped him or her remember the strategy. Consistently included among every response were the amount of practice, verbal rehearsal of strategy steps, and the acronym PIRATES. While these responses were subjective, they are supported by the findings and contentions of previously cited writers as to effective instructional procedures. For example, Bangert-Drown et al. (1983) found a correlation between amount of instruction/practice and the effectiveness of a coaching program. Markel (1981) promoted the use of self-verbalization techniques similar to verbal rehearsal to assist students in acquiring test-taking skills. Finally, it appears that the use of an acronym (e.g., PIRATES) does help students remember the instructional steps (Pressley & Levin, 1982).

There were no apparent differences in acquisition rates between LD and EH subjects. Trials to criterion on verbal rehearsal, cluster probes, and acquisition probes were similar between these two categories of subjects. It appears there is some support that the instructional procedures developed for LD adolescents are effective with EH adolescents as well. The only variable that seemed to affect acquisition rate was I.Q. as measured by the WISC-R. The records for each subject regarding trials to mastery for the verbal rehearsal step show that the two subjects with the lowest I.Q.s (78 and 80) took two or three more trials to obtain the 100 percent mastery level. In this step, subjects were required to name all seven components (e.g., "P--prepare") of PIRATES and describe the behaviors in each

component (e.g., "S--survey. Make sure all questions are answered. Don't change answers unless very sure it is wrong."). Consequently, it may be that students with lower obtained scores on the WISC-R require more trials when memorizing and verbalizing a complex strategy. It should be noted, however, that students with slightly higher I.Q.s (i.e., 81 and 84) did not require more trials.

Lack of motivation to learn is a common descriptor of adolescents with learning problems. Subjects in this study appeared motivated during instructional sessions. Generally, no external reinforcement or verbal prompts were needed to maintain on-task behavior. While a number of variables might account for subject motivation (e.g., small group size, novelty of task or instructor) one aspect of the instructional procedure seemed highly motivating. Students set personal goal dates for each instructional step and charted their probe scores. Self-charting, once taught, was conducted spontaneously, without prompting. Further incidental evidence of the motivational power of this technique lies in the fact several students requested charts to take home in order to show their parents the visual display of their progress.

In summary, subjects did acquire the strategy to the prescribed mastery level. Instructional procedures used were effective in teaching the strategy and little, if any, external motivation techniques were required to maintain subject involvement with the task of acquiring proficiency with PIRATES. Additionally, the amount of time to teach the skill to mastery was six hours spread across four weeks. This seems a reasonable amount of instructional time given the possible benefits to students.

Question 2: After Acquiring a Test-Taking Strategy at 90 Percent Mastery Level, Are Regular Classroom Test Scores of LD and EH Secondary Students Improved?

Results for the second question, regarding generality, are more equivocable than those for the first question of acquisition. In order to clearly show control in single subject designs, baseline data should exhibit low variability and a level or decelerating trend for the effect of the intervention to be readily apparent (Baer, Wolf, & Risley, 1968). In the case of the multiple-baseline across subjects design used in this study, intervention was staggered in order to determine if an obvious change in the data level or trend occurred only when the strategy PIRATES was introduced. Unfortunately, baseline data collected in the present study exhibited relatively large variability, thus lowering stability of the baseline. Several factors may account for variability of the data. These factors include student attendance during presentation of content assessed, students' prior knowledge of content, or the teachers' mastery or interest of the knowledge being tested. Whatever the reason, variability of subject test scores during baseline ranged from 27 to 56 percentage points. When variability of subjects' test scores between the two experimental conditions are compared, it is found the range of test scores decreased a minimum of 10 percentage points for 9 of the 12 subjects. The smallest decrease in range was 10 percent and the largest 38 percent. While arguable, it does appear use of the strategy PIRATES may decrease variability of performance on classroom tests, a contention which supports the idea that test-wiseness is a source of variance in test scores (Millman et al., 1965; Oakland,

1972; Rowley, 1974). The value of reduced variability in test scores may be the increased frequency of passing grades which in turn increases the student's access to natural reinforcers such as parent and teacher approval. In theory at least this in turn should be motivating to the student and success in the mainstream class may begin to assume some intrinsic reward.

As noted earlier, the variability of data made visual inspection of data trends and levels between conditions difficult. In order to make differences in data trends more visible, trend lines were calculated and drawn using a split-middle procedure developed by White and Haring (1980). Analysis of trend lines between conditions was made to detect effects of intervention. Positive effects were identified if there was an accelerating trend following a flat or decreasing baseline or by a flat baseline trend followed by a flat trend at a higher level in the second condition. Using the above criteria, some evidence exists that six of the subjects increased test scores due to their acquisition of the PIRATES strategy. An additional two subjects showed substantial gains initially followed by decaying trends. This could be evidence of lack of generalization maintenance. Due to external constraints such as time, scheduling, and reticence of the school administration to have the subjects "pull out" time minimized, each subject received only one generalization session following acquisition. Lack of consistent prompting and discussion regarding continued use of the strategy may account for the decay in trend.

An implication for instruction is that consistent generalization and maintenance sessions may have maintained the initial improvement. If, in the future, resource teachers were teaching the strategy, they would be in a better situation than the experimenter to provide these prompts more frequently and could remind the student to use the strategy immediately prior to taking a test in a mainstream class.

No apparent reasons for the remaining four subjects' lack of improving trends or higher levels for the during/after condition were found. Analysis of the variables of I.Q., reading level, and acquisition level did not show a consistent pattern to explain non-improving trends. However, three of the four students having non-improving trends were categorized as EH, a possible indication that type of exceptionality may be an influencing variable. A strong conclusion should not be drawn on the relationship between exceptionality though due to the small number of subjects.

An additional method of analyzing differences between conditions was used in this study. Test score means and medians within each condition were calculated and then compared to see what difference, if any, occurred. Results showed all subjects, with the exception of one, increased in both mean and median scores across conditions. Additionally, when the more conservative figure (mean or median) was used, nine of the subjects showed increases. When mean and median scores were transformed to letter grades, nine of the subjects increased at least one grade between the two conditions.

Using descriptive statistics, such as the mean and median, to compare differences across experimental conditions is not the ideal

procedure in single subject design. However, due to the nature of the variable being measured (test scores) and the variability of the data, descriptive statistics were deemed appropriate. While the strong evidence of control offered by the multiple-baseline design was in part sacrificed, nine replications of grade increases between conditions does offer some support of the efficacy of the PIRATES strategy.

Up until now the issue of generalization has been discussed in terms of effect (e.g., trends, increases in mean and median between conditions) and, while these are critical to the issue of efficacy, other more subjective measures of generalization exist. Periodically, tests in subjects' mainstream classes were examined for physical evidence of the PIRATES strategy. Physical evidence sought included the writing of PIRATES on the test form, underlining key words, checks indicating items were skipped if the subject did not know the answer, and answering all questions. Most subjects' tests did show they were indeed using all or part of the strategy. Examination of the tests also revealed the majority of the tests were objective and contained predominantly multiple-choice and true-false items.

In summary, the results of this study, while equivocable, do provide evidence that secondary EH and LD students can learn a complex test-taking strategy to a high level of mastery and apply it to tests taken in mainstream classrooms. While control over the dependent measure (test scores) was not clearly established, due largely to the natural variability of the data, there was evidence that test scores

and letter grades did increase for most subjects after strategy acquisition. Relatedly, the acquisition and use of test-taking strategies such as the one used in this study generally do not show large impact on test scores. However, even relatively small changes in scores (e.g., 10 percent) can have a large functional impact such as changing a failing grade to a passing one.

The results reported in this study are similar to those of other studies (Callenbach, 1973; Eakins et al., 1976; Jongsma & Warshauer, 1975; Lee & Alley, 1981), however comparison must be limited since all but one study (Lee & Alley, 1981) used scores on standardized achievement tests as the dependent measure. Lee and Alley used a regular classroom score as a posttest measure but compared it to an experimenter-made pretest, a marked difference from the present study which used classroom tests as the sole dependent variable. Also, this study attempted to examine effects over time while the Lee and Alley study used a single posttest measure leaving the question of durability of effect unanswered.

The standard for academic success in mainstream classrooms is, and most likely will continue to be, scores on tests. The present study adds with some equivocation to the body of evidence that test-wiseness or the existence or absence of test-taking skills is a source of variance in test scores. Since evidence exists that many mildly handicapped students have not acquired some of these skills, it seems important to teach them and teach them well. Acquiring test-taking skills may teach students to use more productive, systematic behaviors in test situations than impulsivity, distractibility, or just giving up.

Limitations

Certain characteristics of design and procedures used in this study limit extrapolation and interpretation of results. First, the study used as a dependent variable grades on regular classroom tests. No conclusions should be made regarding the effectiveness of the PIRATES strategy with other types of tests such as standardized achievement tests or intellectual tests. Relatedly, the skills included in the test-taking strategy were developed for taking objective tests as opposed to essay type tests.

The small number of subjects used in this study, while appropriate for the design, limit conclusions as to how effective the strategy may be with other, similar students. Finally, as discussed previously, the relatively large variability of data decreases the amount of apparent control of subject behavior.

Future Directions

Based on the findings and limitations of the present study as well as related investigations, the suggestions for future investigations regarding test-taking skills are as follows:

1. Considering that the present study is, to date, the only study utilizing classroom test scores as the sole dependent variable, more studies of this type are needed. Test scores determine success in the mainstream classroom and school in general. Therefore, the only true measure of the generality of test-taking skills is whether they do indeed result in increased test scores.

2. Further investigations into the efficacy of the PIRATES strategy should be made. These replications should include a larger number of students and be based in secondary or postsecondary settings. If possible, future studies should collect test scores from more than one mainstream class in order to investigate this aspect of generalizability. The present study used scores from one mainstream class per subject since the majority of subjects were only enrolled in one.

3. The use of teachers to instruct students in PIRATES rather than the experimenter would make the design more realistic and possibly increase confidence in the worthiness of the technique.

4. It would be valuable to investigate the efficacy of each component of PIRATES in isolation. It may be that some components are more useful than others and the strategy could be shortened. This would probably then reduce the amount of time necessary for direct instruction and increase the practicality of the strategy for use in public school settings.

5. Lastly, investigating the effects of combining PIRATES with study skill strategies may provide teachers with a learning strategy package designed to increase performance on classroom tests even more.

REFERENCES

Alberto, P. A., & Troutman, A. C. (1982). Applied behavior analysis for teachers. Columbus, OH: Charles E. Merrill.

Alley, G. R., & Deshler, D. D. (1979). Teaching the learning disabled adolescent: Strategies and methods. Denver: Love.

Alley, G. R., Deshler, D. D., & Warner, M. M. (1979). Identification of LD adolescents: A Bayesian approach. Learning Disabilities Quarterly, 2, 76-83.

Armbruster, B. B., Echols, C. H., & Brown, A. L. (1983). The role of metacognition in reading to learn: A developmental perspective (Reading Education Report No. 40). Champaign, IL: Center for the Study of Reading.

Baer, D. M. (1979). Generalization. Paper presented at the Human Development and Family Life Professional Seminar, Lawrence, KS.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1(1), 91-97.

Bangert-Drown, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53(4), 571-585.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. Educational Researcher, February, 14-21.

Brown, D. (1982). Increasing test-wiseness in children. Elementary School Guidance and Counseling, 16(2), 180-186.

Callenbach, C. (1973). The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second grade students. Journal of Educational Measurement, 10, 25-30.

Carman, R. A., & Adams, W. R. (1972). Study skills: A student's guide for survival. New York: Wiley.

Cohen, S. A., & Foreman, D. I. (1978). Scoring high in reading. New York: Random House School Division.

Costar, E. (1980). Scoring high in reading: The effectiveness of teaching achievement test-taking behaviors. Elementary School Guidance and Counseling, 15, 157-159.

Crehan, K. D., Koehler, R. A., & Slakter, M. J. (1974). Longitudinal studies on test-wiseness. Journal of Educational Measurement, 11, 209-212.

Cuthbertson, E. B. (1979). An analysis of secondary testing and grading procedures. Unpublished master's thesis, University of Kansas, Lawrence.

Dansereau, D. (1978). The development of a learning strategies curriculum. In H. F. O'Neill (Ed.), Learning strategies (pp. 1-29). New York: Academic Press.

Deno, E. (1970). Special education as developmental capital. Exceptional Children, 37, 229-237.

Deshler, D. D., Alley, G. R., Warner, M. M., & Schumaker, J. B. (1981). Instructional practices for promoting skill acquisition and generalization in severely learning disabled adolescents. Learning Disabilities Quarterly, 4, 415-421.

Deshler, D., & Schumaker, J. B. (in press). Intervention issues related to the education of LD adolescents. Journal of Learning Disabilities.

Eakins, D. J., Green, D. S., & Bushnell, D. (1976). The effects of an instructional test-taking unit on achievement test scores. Journal of Educational Research, 70, 67-71.

Epstein, M. H., & Cullinan, D. (1984). Research issues in behavior disorders: A national survey. Behavioral Disorders, 10, 56-59.

Flavell, J. (1977). Cognitive development. Englewood Cliffs, NJ: Prentice-Hall.

Florida Department of Education. (1981). Program manual: Emotionally handicapped (Vol. II-E). Tallahassee: Bureau of Education for Exceptional Students.

Florida Department of Education. (1981). Program manual: Specific learning disabilities (Vol. II-F). Tallahassee: Bureau of Education for Exceptional Students.

Fluitt, J. L., & Gifford, C. S. (1980). Who's teaching teachers how to teach test-wiseness? Contemporary Education, 51, 152-154.

Forness, S. R., & Dvorak, R. (1982). Effects of test time limits on achievement scores of behavior disordered adolescents. Behavioral Disorders, 7(4), 207-212.

Friend, M., & McNutt, G. (1984). Resource room programs: Where are we now? Exceptional Children, 51(2), 150-155.

Fueyo, V. (1977). Training test-taking skills. Psychology in the Schools, 14(2), 180-184.

Gifford, C. S., & Fluitt, S. L. (1978). Programs to improve test-taking skills. Phi Delta Kappan, 60(4), 318-322.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). The fourth R: Research in the classroom. Columbus, OH: Charles E. Merrill.

Havertape, J. F. (1976). The communication functions in learning disabled adolescents: A study of verbalized self-instructions (Doctoral dissertation, University of Arizona, 1976). Dissertation Abstracts International, 37/03A, 1489.

Heller, H. W. (1981). Secondary education for handicapped students: In search of a solution. Exceptional Children, 47, 582-583.

Hughes, C. A., Ruhl, K. L., & Korinek, L. (in press). Predictive research on classroom interventions used with emotionally handicapped adolescents. Programming for adolescents with behavior problems. Reston, VA: Council for Children with Behavior Disorders.

Jongsma, L. A., & Warshauer, T. (1975). The effects of instruction in test-taking skills upon student performance on standardized tests. New Orleans, LA: New Orleans University. (ERIC Document Reproduction Service No. ED 114408)

Kauffman, J. M. (1981). Characteristics of children's behavior disorders. Columbus, OH: Charles E. Merrill.

Keogh, B. (1971). Hyperactivity and learning disorders: Review and speculation. Exceptional Children, 38, 101-110.

Kuhn, D. (1974). Inducing development experimentally: Comments on a research paradigm. Developmental Psychology, 10, 590-600.

Lakin, C. K. (1983). Research-based knowledge and professional practices in special education for emotionally disturbed students. Behavioral Disorders, 8(2), 128-137.

Lange, B. (1981). Promoting test-wiseness. Journal of Reading, 24, 740-743.

Lee, P., & Alley, G. R. (1981). Training junior high school LD students to use a test-taking strategy (Research Report No. 38). Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Link, D. B. (1980). Essential learning skills and the low achieving student at the secondary level: A rating of the importance of 24 academic abilities. Unpublished master's thesis, University of Kansas, Lawrence.

Lue, M. S. (1981). The use of instructional arrangements and their relationship to on-task behaviors with learning disabled, emotionally handicapped, and regular class children (Doctoral dissertation, University of Florida, 1981). Dissertation Abstracts International, 42, 2614.

Markel, G. (1981). Improving test-taking skills of LD adolescents. Academic Therapy, 16, 333-342.

McPhail, I. P. (1978). Why teach test-wiseness. Journal of Reading, 25(1), 32-38.

Mercer, C. D., Bott, D. A., & O'Shea, L. J. (1984). A consumer's guide to basic practices in group and single subject research. Unpublished manuscript, University of Florida, Gainesville, FL.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Oakland, T. (1972). Effects of test-wiseness materials. Journal of School Psychology, 10, 355-360.

Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The mnemonic keyword method. Review of Educational Research, 52, 61-91.

Rowley, G. (1974). Which examinees are most favoured by the use of multiple choice tests? Journal of Educational Measurement, 11, 15-23.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive domain. Review of Educational Research, 49(2), 252-279.

Schumaker, J. B., & Clark, F. L. (1982). An approach to learning strategies training for groups of secondary students (Research Monograph No. 11). Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Schumaker, J. B., & Deshler, D. D. (1983). Setting demand variables: A major factor in program planning for the LD adolescent. Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Schumaker, J. B., Deshler, D. D., Alley, G. R., & Warner, M. M. (1983). Toward the development of an intervention model for learning disabled adolescents: The University of Kansas Institute. Exceptional Education Quarterly, 4(1), 45-75.

Schumaker, J. B., Deshler, D. D., Denton, P., Alley, G. R., Clark, F. L., & Warner, M. M. (1981). Multipass: A learning strategy for improving reading comprehension (Research Report No. 33). Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Schumaker, J. B., Deshler, D. D., Nolan, S., Clark, F. L., Alley, G. R., & Warner, M. M. (1981). Error monitoring: A learning strategy for improving academic performance of LD adolescents (Research Report No. 38). Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Sidman, M., & Willson-Morris, M. (1974). Reading test and stimulus control. Journal of Applied Behavior Analysis, 7, 327-332.

Skrtic, T. M. (1980). The regular classroom interactions of learning disabled adolescents and their teachers (Research Report No. 8). Lawrence, KS: The University of Kansas, Institute for Research in Learning Disabilities.

Slakter, M. J., Koehler, R. A., & Hampton, S. H. (1970). Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 7, 119-122.

Smith, D. D., Deshler, D. D., Hallahan, D., Lovitt, T., Robinson, S., Voress, J., & Ysseldyke, J. (1984). Minimum standards for the description of subjects in learning disabilities research reports. Learning Disability Quarterly, 7(3), 221-225.

Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple choice items. Journal of Educational Measurement, 19(2), 211-220.

Stevens, R., & Rosenshine, B. (1981). Advances in research on teaching. Exceptional Education Quarterly, 2, 1-9.

Strain, P. S., McConnell, S., & Codisco, L. (1983). Special educators as single-subject researchers. Exceptional Education Quarterly, 4, 40-51.

Swift, M. S., & Spivak, G. (1969). Clarifying the relationship between academic success and overt classroom behavior. Exceptional Children, 39, 392-399.

Tawney, J. W., & Gast, D. L. (1984). Single subject research in special education. Columbus, OH: Charles E. Merrill.

Warner, M. M., Schumaker, J. B., Alley, G. R., & Deshler, D. D. (1980). Learning disabled adolescents in the public schools: Are they different from other low achievers? Exceptional Education Quarterly, 1, 27-36.

White, O. R., & Haring, N. G. (1980). Exceptional teaching. Columbus, OH: Charles E. Merrill.

APPENDIX A
INSTRUCTIONAL PROCEDURES

Step 1: Pretest

The purpose of this step was to provide a baseline for the subjects' use of test-taking skills contained in the PIRATES strategy. Probe sheets resembling regular classroom tests and providing opportunities to use each component of PIRATES were used. Administration and evaluation of the pretest probes was implemented in the following sequence:

1. A pretest probe was distributed to subjects along with an explanation that the purpose for the test is to discover how well they take tests. The subjects were told that some questions were easy and some they would not know the answer to because some questions were about made up words or concepts, and not to be worried, just do the best they could. Additionally, they were told that performance on the test would not count for any grades in their classes and no one else other than the experimenter would know how they did. Any questions by the subjects were answered that did not provide direct information regarding specific test-taking strategies.

2. Performance on pretest probes was based on the ratio of test-taking skills performed and the opportunity to perform them. A percentage was calculated from the ratio and recorded for each subject.

3. After the appropriate number of pretest probes were administered and scored, test results were communicated to the subjects individually and their current test-taking behaviors, as demonstrated on the probes, were described to them.

Step 2: Describe

The purpose of this step was to describe the steps in the PIRATES strategy and to provide a rationale for using the strategy. When conducting this session(s), the following instructional sequence was used:

1. Introduce and describe the concepts of strategies and test-wiseness, and how they apply to taking tests. Explain that they will be learning a test-taking strategy called PIRATES that will help them score higher on tests by teaching them to be test-wise.

2. Explain how using the strategy will help them score higher on tests and provide examples of how students have scored higher after learning test-taking strategies.

3. Describe each component of PIRATES briefly, providing examples when appropriate.

4. Summarize the lesson and solicit questions.

Step 3: Model

The purpose of this step was to demonstrate, while verbally describing, the PIRATES strategy in its entirety. When conducting this session(s), the following instructional sequence was used:

1. Prior to actual modeling, a short review of the previous step was given and a verbal advance organizer was provided.

2. Demonstrate, while verbally describing the behavioral sequence of PIRATES. This was done on an overhead projector using a transparency of a sample pretest probe.

3. After modeling the strategy, subjects were asked if they had any questions.

Step 4: Verbal Rehearsal

The purpose of this step was to allow subjects to verbally rehearse steps of the PIRATES strategy to a performance criterion of 100 percent correct. In addition to recalling the name of the strategy step (e.g., "P" means prepare) subjects had to verbalize main aspects of the step (e.g., "I" means instructions. First read carefully and all the way through, then mark the key words).

When conducting verbal rehearsal sessions, the experimenter used the following teaching sequence:

1. Review the strategy steps by asking the students to name them and their key components. Subjects may use cues provided on an overhead transparency to answer.

2. Provide an advance organizer explaining that the purpose of the session is to memorize the steps and that it will be important to be able to do this when using the strategy in an actual test situation.

3. Use a "rapid fire" verbal rehearsal technique whereby the subjects name the steps and components as fast as they can. Initially, the steps will be displayed on an overhead projector, but after several rounds, the projector will be turned off.

4. Time will be allotted for individual rehearsal or in pairs.

5. When a subject has indicated that he/she is ready, they will privately orally list the steps and components to the instructor.

6. When subjects have demonstrated 100 percent correct performance, they will then begin to practice using all the steps of the PIRATES strategy.

Step 5: Practice and Feedback in Controlled Materials

The purpose of this step was to provide practice in "clusters" of the behaviors included in the PIRATES strategy. For instructional purposes, the steps of this strategy were grouped into two clusters as follows: Cluster 1, steps 1-9; Cluster 2, steps 10-12. Each cluster was practiced individually in numerical sequence. Subjects demonstrated proficiency in a cluster before instruction and practice in the next cluster began. Administration and evaluation of cluster probes was implemented in the following sequence:

1. Review the PIRATES strategy and have subjects name the steps and components of the strategy. A cue card with strategy acronym and names of the component was used by the subjects to assist them.

2. Provide a review of the steps contained in the first cluster.

3. Provide a probe for the first cluster and explain that this sheet will allow them to practice the steps and permit the experimenter to see how well they are using the steps in the cluster.

5. Initially, monitor practice sessions and give immediate help and corrective feedback.

6. Score and record the performance of each subject on the cluster probe.

7. At the beginning of the next practice session provide feedback on performance on the previous cluster probe.

8. When subjects have demonstrated proficiency of the steps in the first cluster, the next cluster will be introduced with practice following the same sequence described above.

9. When subjects have demonstrated proficiency in all three clusters, they will proceed to the next step.

Step 6: Practice and Feedback in Content Materials

The purpose of this step was to provide practice incorporating all components of the PIRATES strategy. Probe sheets resembling regular classroom tests, providing practice in all components of the strategy were used. Additionally, these probes which were equivalent to test probes administered in the first instructional step (Pretest), were used to assess whether the subjects had acquired the use of the PIRATES strategy.

When conducting the above practice/assessment sessions, the experimenter used the following teaching sequence:

1. Review the PIRATES strategy and have the subjects name the steps and components of the strategy.

2. Provide a verbal advance organizer by stating that the subjects are expected to use everything they have been taught during the unit, and that the sheets will allow them to practice their new skill and allow the experimenter to see how well they know it.

3. Monitor the subjects' practice and give immediate help or corrective feedback.

4. Evaluate and score subjects' performance on the probes by analyzing the probe for evidence indicating correct use of the various components of the strategy.

5. Provide feedback to the subjects individually on their performance. Feedback will be both positive and corrective. The subjects will also be asked to verbalize the correct use of a strategy component omitted or incorrectly used.

6. Decide when each subject has acquired the strategy at a proficient level. When subjects have demonstrated, unassisted, proficient use of PIRATES, generalization procedures will be implemented following the next unit test taken in the regular mainstream classroom.

Step 7: Promoting Generalization

The purpose of this step was to systematically promote generalization of the PIRATES strategy to regular classroom testing situations. This step has two components, orientation and activation, and are described as follows:

A. Orientation

1. Have subjects make cue cards with acronym PIRATES to carry with them.
2. Discuss various forms of tests and how PIRATES can be used or adjusted when taking them. Samples of actual regular class tests will be shown on the overhead to facilitate this discussion. At the end of this session(s), subjects will be able to verbalize how PIRATES can be adjusted to use with different test formats.

B. Activation

1. Tell the subjects to use PIRATES when taking their next unit test. After they have taken a unit test, a copy will be obtained and analyzed by the experimenter. After this analysis each subject will meet individually with the experimenter for positive and correct feedback.

APPENDIX B
METHODS FOR ESTIMATING ANSWERS

The following are descriptions of elements or available cues dependent upon construction, that subjects were taught to utilize when estimating answers.

1. Length of option/most qualified option: When guessing, the longer or most qualified option may be the most appropriate guess.

2. Specific determiners: Multiple choice options which contain words such as never or always (see Appendix D) are often incorrect. True-False statements which include specific determiners should be marked false if the test taker does not know the correct answer.

3. Similar options: If two options are the same, but differently worded they should both be assumed to be incorrect.

APPENDIX C
KEY WORDS/PHRASES IN INSTRUCTION

The following words or phrases occur frequently in directions given on classroom test forms (Cuthbertson, 1979). These words and phrases were taught to the subjects as part of the "Instructions" component of PIRATES (behavioral sequence, number four). The words included below that are in parentheses are included to provide context to the key words listed and underlined as follows:

1. (Choose the) best response
2. Indicate (your) choice
3. (Most) appropriate (response)
4. (In the) blank provided
5. Select (the choice)
6. (By) circling (the letter)
7. (Circle the letter) preceding
8. Mark
9. Column
10. Match

APPENDIX D
SPECIFIC DETERMINERS

Options containing specific determiners should be considered incorrect when subjects are estimating answers. The following is a list of words that are specific determiners and are included in the PIRATES strategy:

no

all

never

every

only

always

none

APPENDIX E
PRETEST

Write your name in the upper left hand corner of the test form. Write the date and class period in the lower right hand corner.

A. Indicate the best response by circling the letter preceding it.

1. The Argualan colonists fought the king because of

   a. land rights
   b. cruelty
   c. unfair taxation of export products
   d. water rights

2. The number of protons in krenium is

   a. the same as menium
   b. less than menium
   c. equal to menium
   d. more than kryptium

3. Most people buy food at the

   a. bank
   b. laundry
   c. garage
   d. grocery store

4. Working conditions during the colonization period were

   a. always fair
   b. never fair
   c. usually poor
   d. never discussed

5. A score in football is called

   a. home run
   b. basket
   c. touchdown
   d. dunk

6. The atomic weight of the element krypton is

   a. greater than radium
   b. the same as radium
   c. more than radium
   d. less than radium

B. In the blank provided, write the letter of the most appropriate response.

_____ 1. When the preservative monoglate is added to food

   a. the food never spoils
   b. the food always turns colors
   c. the food never turns colors
   d. the food rarely spoils

_____ 2. The most common use of bason solutions is

   a. fermentation
   b. pigmentation
   c. as a cleaning compound
   d. as the organic catalyst in oil solvents

_____ 3. The United States was interested in purchasing the island territories

   a. for territorial expansion
   b. for tourism purposes
   c. in order to build a strategic air defense base
   d. to obtain its sugar crop

_____ 4. Architects stopped using argronimum in construction because

   a. it was too heavy
   b. it was too expensive
   c. it wasn't strong enough
   d. it weighed too much

_____ 5. Christmas occurs in the month of

   a. September
   b. January
   c. December
   d. November

_____ 6. Anroid compounds

   a. are never used in foods
   b. are always used in canned foods
   c. are only used in cold climates
   d. are usually used in food dye

C. Write T by the statements that are true and F by the statements that are false.

1. Prisoners in federal prisons are usually allowed to use library facilities.

2. Low levels of monoplast in the blood always indicates endematic infection.

3. Every southern farmer who raised silage crops during the 1884 drought lost money.

4. All doctors agree that aspirin is the best treatment for the common form of phrenitis.

5. Many of the countries in the lower peninsula region depend on irrigation for crop watering.

6. If sendium is added to phosphorous it never fails to ignite.

7. None of the farmers in the early 1600's grew more than one kind of crop.

8. Most cancerous cells contain a form of skotema.

9. Radio thermography is seldom used to diagnose apranoma.

10. Few species of spotted clownfish inhabit the southern reefs.

11. The Scari tribe often invaded neighboring tribes to steal horses.

12. Only people who study hard got good grades on the tests.

D. Underline the letter of the best answer.

1. People who are in charge of a sports team are called

   a. waitresses
   b. coaches
   c. congressmen
   d. pilots

2. A person who practices medicine is called a

   a. nurse
   b. teacher
   c. doctor
   d. policeman

3. Most mamtoids

   a. live in warm water
   b. are canabalistic
   c. use coral reefs for building their homes
   d. live in the sand

4. The average South American Colluskian Indian

   a. hunts with a firearm
   b. hunts with a bow and arrow
   c. hunts with a sling shot
   d. hunts with a gun

5. The chemical fertilizer natrite is

   a. usually used on fruit trees
   b. never used on fruit trees
   c. toxic to some plants
   d. toxic to no plants

TEST TAKING SCORE SHEET

Student's Name _____ Date _____

Pretest _____ Posttest _____

Component	Skill	Possible Score	Student Score
Prepare	• Wrote PIRATES on test form	1	
	• Followed directions (name, etc.)	3	
	• Started within 2 minutes	1	
Instructions	• Underlined key words/phrases (how and where to indicate answers)		
	- Section A	2	
	- Section B	2	
	- Section C	4	
	- Section D	2	
	• Indicates answers correctly		
	- Section A	1	
	- Section B	1	
	- Section C	1	
	- Section D	1	
Read	• Eliminated (crossed out) obviously wrong responses	11	
Answer or Abandon	• Abandoned and marked appropriately "no answer" items in each section	4	
Turn Back	• No meausre		
Estimate	• Used effective guessing methods		
	- longest choice	4	
	- similar choices	4	
	- absolutes/non-absolute	16	
Survey	• Answered all questions		
	- Section A	1	
	- Section B	1	
	- Section C	1	
	- Section D	1	
	Total Score	62	

Calculating the Score

$$\frac{\text{Student Score}}{\text{62 (total score)}} = \boxed{}\%$$

APPENDIX F
ACQUISITION SCORES FOR ALL SUBJECTS

Subject	Baseline	Cluster 1	Cluster 2	Acquisition	Maintenance
1	36,33	62,100	88,98	82,86,95	90,90,88
2	33,29,28	64,100	96,100	92,97	77,90
3	36,25,32,31	95,100	94,94	89,92	88
4	33,35	80,93	92,100	94,100	88,76,78
5	34,28,29	66,90	82,95	84,90	84,92
6	22,26,26,23	62,90	82,82	86,86	82
7	36,29	82,97	64,100	95,97	84,82,84
8	36,34,30	70,100	82,95	78,100	90,86
9	33,22,35,32	92,91	89,90	94,100	100
10	34,29	92,96	88,92	90,98	94,80,84
11	32,33,35	78,88	92,94	92,88	84,92
12	38,24,32,33	70,98	75,92	94,100	98

APPENDIX G
GENERALIZATION SCORES FOR ALL SUBJECTS

Subject	Baseline	During Acquisition	After Acquisition
1	64,53,80,65	72,85	44,78,75,76,75,70
2	48,68,48,68,40,60	100	80,40,80,80,50,64
3	48,43,70,60,68,56,43,45	71,58	60,70,65
4	55,82,36,70,59	50,69	90,81,79,65,80,60,80,80,96
5	56,35,32,68,68,80,45	84,86	80,56,78,56,68
6	48,30,35,56,52,80,52,50	50,82,84	65,80
7	50,30,66,50,72	65,80	36,63,76,45
8	64,70,58,60,30,50	55,69	72
9	61,54,89,57,67,58,81,72,56	81,83	78,88,77
10	70,72,36,57,60,50	55,65	53,48,62,68
11	56,60,67,31,85,48,62	55,92	60,77,69,61,64
12	60,76,56,32,68,80,60,60	80	70,75,70

BIOGRAPHICAL SKETCH

Born January 23, 1948, in Memphis, Tennessee, Charles Hughes spent his first seven years in a variety of places as is usual for a military family. After his father retired from the Navy, the Hughes family moved to Novato, California. Mr. Hughes received his B.A. in sociology from The Chico State University in 1970.

In 1971, Mr. Hughes moved to Louisville, Kentucky, where he was a teacher of the trainable mentally handicapped. At this time he enrolled at Eastern Kentucky University majoring in special education. Upon receiving his M.A.Ed. in special education in 1974, Mr. Hughes taught emotionally handicapped adolescents for three years in the Louisville, Kentucky, schools. In 1977, Mr. Hughes moved to Charlotte, North Carolina, where he was a consultant to emotionally handicapped programs for the State Department, Division of Exceptional Children.

Mr. Hughes, his wife Kathy Ruhl, and daughter Hilary moved to Gainesville, Florida, in 1981, for the purpose of obtaining a Ph.D. in special education from the University of Florida. During this period Mr. Hughes worked full-time for two years as a consultant and educational diagnostician for the Multidisciplinary Diagnostic

and Training Program and attended classes part-time. At the completion of his doctoral program, Mr. Hughes will take a teaching and research position with the Special Education and Communication Disorders Department at The Pennsylvania State University.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Cecil D. Mercer, Chairman
Professor of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Myron A. Cunningham
Professor Emeritus of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Lawrence J. O'Shea
Assistant Professor of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Cary L. Reichard
Professor of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stuart E. Schwartz
Professor of Special Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward C. Turner
Associate Professor of Instruction and Curriculum

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1985

David C. Smith (N)
Dean, College of Education

Dean, Graduate School